



ISAB Review of Spring Chinook Salmon and Steelhead Models for Assessing Responses to EIS Alternatives for the Willamette Valley System

INDEPENDENT SCIENTIFIC ADVISORY BOARD
ISAB 2023-1 MARCH 3, 2023 (UPDATED MAY 22)

Update Statement - May 22, 2023:

The original version of this ISAB Willamette Models Review Report ([ISAB 2023-1](#), March 3, 2023) was updated on May 22, 2023. Statements in the text were revised to make it clear that UBC and NOAA models used historical years and the EDT and OSU models used three specific years to represent differences (dry, normal, wet) in hydrology. A text box at the top of Figure 3 was changed from referring to three years “2011, 2015, and 2016” to “NAA and alternatives.” In addition, references to the OSU models as “secondary” models were removed, and instead the OSU models are categorized as one of the four “primary models,” along with the UBC, NOAA, and EDT primary models. This editorial change makes the categorization of “primary” models consistent throughout the report. (See the [list of specific revisions](#).)

Cover design by Eric Schrepel, Technical and Web Data Specialist, Northwest Power and Conservation Council

Photos clockwise from top left: Cougar Reservoir upstream of dam (Erik Merrill), Detroit Dam (photo courtesy of the U.S. Army Corps of Engineers, Greg Taylor), Minto Fish Collection Facility, North Santiam River (U.S. Army Corps of Engineers, Scott Clemans), and McKenzie River (Tony Grover).



Independent Scientific Advisory Board

for the Northwest Power and Conservation Council,
Columbia River Basin Indian Tribes,
and National Marine Fisheries Service
851 SW 6th Avenue, Suite 1100
Portland, Oregon 97204

ISAB Members – Authors and Contributors

Courtney Carothers, Ph.D., Professor, Department of Fisheries, College of Fisheries and Ocean Sciences,
University of Alaska Fairbanks

John Epifanio, Ph.D., (ISAB Vice-Chair) Former Principal Scientist with the Illinois Natural History Survey and
Research Professor with the University of Illinois; now based in Portland, Oregon

Stanley Gregory, Ph.D., (ISAB Chair) Professor Emeritus, Department of Fisheries, Wildlife, and Conservation
Sciences, Oregon State University, Corvallis, Oregon

Dana Infante, Ph.D., Associate Professor and Associate Chair of Research in the Department of Fisheries and
Wildlife, Michigan State University

James Irvine, Ph.D., Emeritus Research Scientist, Fisheries and Oceans Canada, Pacific Biological Station,
Nanaimo, British Columbia

Yolanda Morbey, Ph.D., Associate Professor, Department of Biology, Western University, Ontario, Canada

Thomas P. Quinn, Ph.D., Professor, School of Aquatic and Fishery Sciences, University of Washington, Seattle

Kenneth Rose, Ph.D., France-Merrick Professor in Sustainable Ecosystem Restoration at Horn Point
Laboratory of the University of Maryland Center for Environmental Science

Carl Schwarz, Ph.D., Professor Emeritus of Statistics and Actuarial Science at Simon Fraser University, Canada
(ad hoc ISAB member)

Desiree Tullos, Ph.D., P.E., Professor at Oregon State University, Biological and Ecological Engineering
Department

Thomas Wainwright, Ph.D., Retired Research Fishery Biologist, NOAA Fisheries; now based in Bend, Oregon

Ellen Wohl, Ph.D., Professor of Geology and University Distinguished Professor, Department of Geosciences,
Colorado State University, Fort Collins

Michael Young, Ph.D., Emeritus Research Fisheries Biologist, Rocky Mountain Research Station, U.S. Forest
Service, Missoula, Montana

ISAB Ex Officios and Manager

Michael Ford, Ph.D., Director of the Conservation Biology Program, Northwest Fisheries Science Center,
Seattle, Washington

Leslie Bach, Ph.D., Senior Program Manager, Northwest Power and Conservation Council, Portland, Oregon

Robert Lessard, Ph.D., Fisheries Scientist, Columbia River Inter-Tribal Fish Commission, Portland, Oregon

Erik Merrill, J.D., Independent Science Manager, Northwest Power and Conservation Council, Portland,
Oregon

ISAB Review of Spring Chinook Salmon and Steelhead Models for Assessing Responses to EIS Alternatives for the Willamette Valley System

Contents

| | |
|---|----|
| List of Acronyms and Abbreviations | iv |
| Acknowledgements..... | v |
| Executive Summary..... | vi |
| I. Introduction | 1 |
| A. Review Request | 1 |
| B. Willamette Valley System and EIS Background | 1 |
| C. Review Process | 3 |
| D. Report Organization | 6 |
| II. ISAB Comments on the Use of Multiple Models to Assess ESA-listed Fish Responses to Willamette Valley System EIS Alternatives..... | 7 |
| A. Background..... | 7 |
| B. Multiple Models used by the Corps for the Willamette Draft EIS..... | 9 |
| C. ISAB Comments..... | 14 |
| 1. Model Selection and Differences..... | 14 |
| 2. Configuration and Shared Information..... | 15 |
| 3. Independence of Predictions | 16 |
| 4. Combining Results..... | 17 |
| D. Going Forward | 24 |
| III. ISAB Review Comments on the Individual Models..... | 26 |
| A. Fish Benefits Workbook..... | 26 |
| B. Review of the University of British Columbia Models | 28 |
| 1. Summary Comments..... | 28 |
| 2. ISAB Answers to the Review Questions | 29 |
| 3. ISAB Comments on UBC Report Chapters | 34 |
| C. Review of the NOAA Models | 66 |
| 1. Summary Comments..... | 66 |

| | |
|--|-----|
| 2. ISAB Answers to Review Questions | 67 |
| 3. Recommended Changes throughout the Document..... | 73 |
| 4. ISAB Comments on NOAA Report Chapters | 73 |
| D. Review of Ecosystem Diagnosis and Treatment Models..... | 96 |
| 1. Summary Comments..... | 96 |
| 2. Background: EDT | 96 |
| 3. ISAB Answers to Review Questions | 99 |
| 4. Editorial Comments | 109 |
| E. Review of OSU Fish Survival and Flow Models | 111 |
| 1. Summary Comments..... | 111 |
| 2. Background: OSU Fish Survival and Flow Models..... | 112 |
| 3. ISAB Answers to Review Questions | 115 |
| Appendix 1: ISAB Review Questions for the Individual Models and the Multi-model Approach | 131 |
| A. Questions for Individual Models | 131 |
| B. Questions for the ISAB’s recommendations to the Corps about using multiple models... | 133 |
| References | 135 |

List of Figures

| | |
|--|----|
| Figure 1. The Willamette River Basin, showing Willamette Valley System projects, including dams and reservoirs, adult fish collection facilities, and fish hatcheries. | 2 |
| Figure 2. Exchange of information between the supporting models (grey boxes) and the primary models (white boxes) in the Corps analysis for the WVS EIS. | 10 |
| Figure 3. Flowchart showing the steps in the overall synthesis of the multiple models. | 13 |
| Figure 4. Graphical display of the Meets/Fails tables for the three metrics (Abundance, Productivity, and Risk) for the NOAA and the UBC models, and the resulting adverse effects rating of Major or Minor impacts for ten selected population-alternative combinations reported in the Draft EIS. | 21 |
| Figure 5. Graphical display of the Meets/Fails for the same ten selected population-alternative combinations shown in Figure 4 showing the three metrics (Abundance, Productivity, and Risk) by model. | 22 |

Figure 6. Graphical display of the rankings of wet, dry, and normal years from the EDT model for the same ten selected population-alternative combinations shown in Figure 4..... 23

Figure 7. Prior distributions versus the “true” parameter for the ϕ_{SUJ} simulation in Appendix J. 64

Figure 8. Comparison of piece-wise linear interpolation used in the NOAA report with the TRT step-wise interpolation..... 77

Figure 9. Viability curves showing relationship between risk levels and population persistence categories (example based on Chinook curve)..... 79

Figure 10. Comparison of the estimates of NOR spawners from the EDT, NOAA, and UBC models for the EIS alternatives (Appendix E in draft EIS 2022)..... 100

List of Acronyms and Abbreviations

ABC – approximate Bayesian computation

AIC – Akaike information criterion

ARMA – Autoregressive Moving Average

BOR – Bureau of Reclamation

BUGS – Bayesian inference Using Gibbs Sampling

Corps – U.S. Army Corps of Engineers

CC – climate change

CE-QUAL-W2 - Portland State University hydrodynamic and water quality model

CJS – Cormack-Jolly-Seber model

DPEIS – Draft Programmatic Environmental Impact Statement (herein also Draft EIS)

EDT Models – ICF Ecosystem Diagnosis and Treatment Model

EIS – Environmental Impact Statement

ESA – U.S. Endangered Species Act

FBW – Fish Benefits Workbook

FWP – Northwest Power and Conservation Council’s Fish and Wildlife Program

HOR – Hatchery origin return

infm – informative priors

IPA – Integrated Passage Assessment

JAGS - engine for running BUGS in Unix-based environments

KS - Kolmogorov–Smirnov statistical test

LCM – life-cycle model

NAA – No Action Alternative

NMFS – U.S. National Marine Fisheries Service

NOAA – U.S. National Oceanic and Atmospheric Administration

NOAA Models - NOAA Upper Willamette River Life Cycle Model

NOR – Natural origin return

OSU Models - Oregon State University Flow and Fish Survival Models

PDO – Pacific Decadal Oscillation

PEIS – Programmatic Environmental Impact Statement

pHOS – proportion of hatchery origin spawners

pNOB – proportion of natural origin brood

PRB – percentage relative bias

PSM – prespawning mortality

QA/QC – quality assurance and quality control

QET – quasi-extinction threshold (risk)

ResSim – Reservoir Simulation model

R – R computer language

R/S – recruits per spawner

RSS – relative smolt survival

SARs – smolt to adult ratios

SAS – adult survival rate

SDM – Structured Decision Making

SUJ – Sullivan Juvenile Facility System – Willamette Valley System

SWIFT - Science of Willamette Instream Flows Team

TGD – total dissolved gas

UBC Models - University of British Columbia Integrated Passage Assessment Models

USACE – U.S. Army Corps of Engineers

USFWS – U.S. Fish and Wildlife Service

USGS – U.S. Geological Survey

VSP – Viable Salmonid Populations

WFF – Willamette Falls

WVS – Willamette Valley System

Acknowledgements

Numerous individuals assisted the ISAB with this report. Their help and participation are gratefully acknowledged.

Rich Piaskowski and Rachel Laird, U.S. Army Corps of Engineers, Portland District, were especially helpful in coordinating communication, organizing briefings, and providing information and support from the assignment's development to its completion.

We greatly appreciated the briefings, responses to our requests for information, and constructive approach to review dialogue by the teams and individuals who developed the models and analyses that supported or were the subject of our review:

- University of British Columbia modeling team: Murdoch McAllister, Mairin Deith, Roberto Licandeo, Aaron Greenberg, Eric Parkinson, Oliver Murray, and Tom Porteus
- NOAA-Northwest Fisheries Science Center modeling team: Rich Zabel, Jeff Jorgensen, Morgan Bond, and Jim Myers
- ICF Ecosystem Diagnosis and Treatment modeling team: Laura McMullen, Janel Sobota, and Karl Dickman
- Oregon State University, U.S. Geological Survey, Oregon Cooperative Fish and Wildlife Research Unit, modeling team: Jim Peterson and Jessica Pease
- U.S. Geological Survey, Oregon Water Science Center, temperature modeling (CE-QUAL-W2): Laurel Stratton Garvin
- U.S. Army Corps of Engineers: Norm Buccola (temperature model), Josh Roach (ResSim model), Kelly Wingard (WVS EIS Project Manager), and Kathryn Warner (WVS EIS Technical Lead)

Kevin Malone (consultant) also provided background information to us on the Ecosystem Diagnosis and Treatment model as part of our recent review of the Upper Columbia United Tribes' salmon reintroduction planning documents that proved useful for this Willamette review.

The ISAB Ex Officio members – Mike Ford, Bob Lessard, and Leslie Bach – helped define our review and provided important context.

Executive Summary

The U.S. Army Corps of Engineers (Corps) is preparing a Programmatic Environmental Impact Statement (PEIS) to assess the effects of the continued operations and maintenance of the Willamette Valley System (WVS). The WVS consists of ten high-head dams and three run-of-the-river dams, and their associated reservoirs, operated and maintained by the Corps in the Willamette River basin. In April 2021, the Corps requested that the Independent Scientific Advisory Board (ISAB) review a set of models for their scientific strengths and weaknesses in assessing potential responses of ESA-listed spring Chinook salmon and winter steelhead to alternative management actions. The ISAB focused on the fish response models used for the [Draft Programmatic Environmental Impact Statement \(DPEIS or Draft EIS\) for the Willamette Valley System](#), and did not evaluate the merits of the management alternatives or the Corps' decisions on these alternatives.

Four independent teams developed or updated existing models to assess the responses of spring Chinook salmon and steelhead to the management alternatives. This ISAB report describes individual strengths and weaknesses of the following models and also provides suggestions for their collective use when used in a coordinated way as a multiple model analysis:

1. University of British Columbia Integrated Passage Assessment Models (UBC models)
2. NOAA Upper Willamette River Life Cycle Models (NOAA models)
3. ICF Ecosystem Diagnosis and Treatment Models (EDT models)
4. Oregon State University Flow and Fish Survival Models (OSU models)

These four primary models used outputs from three supporting models as inputs:

1. Fish Benefits Workbook (FBW) for juvenile fish dam passage survival and timing estimates
2. USACE Hydrologic Engineering Center ResSim model for reservoir operations affecting river flows
3. Portland State University Water Quality Research Group's CE-QUAL-W2 hydrodynamic and water quality model for stream temperatures

The supporting models captured differences among management alternatives. The ISAB only looked at one of these supporting models and provides some brief comments on how the FBW supporting model was used in the multiple model analysis.

The Corps used the NOAA, UBC, and EDT models (with inputs from the supporting models) to assess effects of EIS management alternatives on spring Chinook and steelhead in the

Willamette River basin. These models predict various combinations of adult abundance, productivity, spatial structure, and diversity, which are metrics for assessing long-term population stability and persistence when using the Viable Salmonid Populations (VSP) approach to evaluate population status. The Corps used the OSU models to provide information on within-year effects of flow management on survival, but the OSU models were not used to evaluate alternatives in the Draft EIS.

The ISAB commends the Corps and the modeling groups for developing an innovative multi-model approach for assessing spring Chinook salmon and winter steelhead responses to management alternatives as part of preparing the Draft EIS for the Willamette Valley System. Multiple models can characterize uncertainty more accurately than single models and lead to more informed decision-making. By design and as implemented appropriately by the Corps for the WVS EIS process, the predictions of the different models offer alternative plausible representations of system response and are not completely independent from each other. Dependence among the models arises because they were selected to answer the same questions of flow management effects, they were developed with knowledge of the other models, and they purposely shared some common information (same field data and inputs from supporting models). Ultimately, the use of multiple models can provide greater confidence than if just one of the models is used in the analysis.

The effectiveness of using multiple models is influenced by how the results (predictions) are combined across the models and interpreted. The multiple model approach documented by the Corps used a thoughtful but complicated scheme to combine the results that relied on professional judgement to interpret the degree of agreement among multiple model predictions. The ISAB considers the scheme used by the Corps to be appropriate, though a more transparent explanation of the Corp's interpretations would strengthen comprehension by stakeholders and the public. The multiple model approach used by the Corps can provide valuable information on uncertainties that should lead to better decisions with higher confidence.

In addition to any further use for the EIS process, the multiple models can be very useful for future assessments and decisions related to implementation, monitoring design, and adaptive management. As part of this review, the ISAB offers several suggestions for improving the multiple model approach used by the Corps. The Corps has already considered many of these suggestions to various degrees, but these considerations either would benefit from further elaboration or their documentation has not been synthesized. The suggestions for the multiple model approach are in three groups:

1. Documentation of why the models were selected, the similarities and differences of the models to facilitate inter-model comparison, and the past performance of the models in similar situations as the WVS EIS;
2. Graphics and further analyses to help interpret the results, such as a description of how the outputs of the supporting models were transformed for use in the four primary models, why predictions differed, the role of shared supporting models versus structural differences in representations within the four models, and graphical displays of how predictions are combined; and
3. Future activities (e.g., workshops, code maintenance) to refine and improve the multi-model strategy, ensure the approach is adaptable to other applications, and provide mechanisms for the models and codes to remain up-to-date and usable.

The value of a multi-model approach depends on the realism of each of the models used. As part of this review, the ISAB reviewed the four primary models for spring Chinook salmon and steelhead developed by the modeling groups. The individual modeling efforts were documented in separate reports that preceded the multi-model comparison. To provide the review in a timely manner and because this review occurred in parallel with the evolving development and application of the models by the Corps, the ISAB reviewed each of the models as if they were used alone to assess the responses to the management alternatives. Some of the ISAB's comments on the individual models may have been addressed by the Corps in their final multi-model analysis. Reviewing the individual models as standalone applications is useful because each review is complete, how the models are used in the multi-model approach can change, and the models may be used as standalone applications for the WVS or in other similar situations in the future.

Overall, the ISAB determined that the four models used by the Corps for the multi-model analysis include the major processes influencing spring Chinook salmon and steelhead life histories and are scientifically sound. The composition and performance of the individual models largely achieves their intended analytical goals. However, they have some important differences, emphases, and areas of overlap. Consequently, combined use of the primary models provides valuable information for assessing potential responses of spring Chinook salmon and steelhead to the operational alternatives. For example, consistency in predicted outcomes increases confidence in the assessments, and different outcomes identify uncertainties, information gaps, and areas for improvement in the models.

The NOAA and UBC models are full life cycle models that predict abundances of Chinook salmon and steelhead by life stage across multiple generations. The EDT and OSU models estimate abundance based on flow conditions for a single year and do not predict effects of

flow management alternatives on long-term population dynamics. The models appropriately differ in temporal resolution (time step): annual for the UBC and NOAA models, monthly for the EDT models, and seasonal with weekly time steps for the OSU models. The spatial domains for the four models also differ, and this affects what responses can be predicted in response to the operational alternatives. These differences in spatial domain are inherent to what the different models were designed to predict and reflect decisions by the modeling teams. For example, the UBC and NOAA models do not represent the flow conditions and habitat of the mainstem, EDT models include the tributaries and the mainstem Willamette River but not flow-related channel and floodplain changes in the mainstem, and the OSU models include the mainstem river and tributaries above dams. Such differences are logical for each model and must be carefully considered in the interpretation of the results from each individual model, as well as when predictions are combined across models.

The independent model reviews offer many comments about the specifics of each model application. Several cross-cutting issues emerged that apply to most, if not all, of the primary models and how each was used for the Draft EIS. These issues are:

- **Calibration and validation** — The models are calibrated differently, which partially reflects their structural differences, and also reflects how each incorporates the available data. The models either have a long history of development and applications and their parameter values have been vetted over time, or detailed calibration is explored (UBC models). Rigorous validation (comparison to an independent dataset) was limited and should be further investigated for all of the models. The different models evaluated the performance of their submodels to varying degrees, and the modeling summary reports described the analyses for the specific submodels that were evaluated. Additional confirmation of the realism of the complete model (i.e., the combined submodels) would advance the confidence in the individual models. The temporal extent of the available data (number of years) in the WVS for Chinook salmon and the limited data for steelhead permeate the calibration of all of the models, especially the NOAA and UBC life cycle models. More attention to evaluating the models for their site-specificity to the WVS will further improve confidence.
- **Interannual variation** – The common use of three specific years to represent differences (dry, normal, wet) in hydrology for EDT and OSU models is sound. Given the importance of assessing model predictions across hydrological conditions, adding additional years (e.g., 2 to 3 years for each of dry, normal, and wet), as well as expected future conditions under climate change, would enable both within and among year-type comparisons of the alternatives and a more realistic sense of the responses within the subbasin’s fish community. No modeling of potential climate change effects on the

alternatives was performed. Consequently, comparisons based on historical water-years may not reflect performance of the alternatives in the future and the results should be interpreted cautiously.

- **Influence of supporting models** – The sharing of the outputs from the same supporting models as inputs to the four primary models is a thoughtful and reasonable way to ensure that model results can be compared without too much forcing of the primary models to agree because of shared information. Various results reported in the modeling summaries indicate that the FBW likely influences the primary model predictions of differences among the flow management alternatives. Determining the relative influence of the different supporting models, versus the internal representations of the primary models, on the prediction of differences among alternatives is critical to proper interpretation of results. This is best conducted with simulations specifically designed to separate the influence of supporting model inputs versus the internal structural differences for the primary models.
- **Mortality** – Different representations of mortality in the models reflect plausible possibilities of what occurs in nature. Two specific aspects of mortality should be further compared among the models: how density-dependence is formulated and represented in different life stages and the varying influence of marine survival.
- **Model uncertainty and sensitivity analyses** – Various sensitivity analyses (small changes in parameters) are included with the individual models, and some report results with prediction intervals. Additional analyses would further leverage the usefulness of each of the models by identifying why predictions differed among the models. Such analyses should be coordinated among the modeling groups, so they can use a standardized approach to varying parameters and can have uniformity in how they report the variability around predictions. In addition, formal uncertainty analyses that use realistic variation in parameters and statistical resampling of parameter distributions to generate expected variances of outputs would be of value. Results of uncertainty analysis would allow inclusion of variability in predictions due to internal model dynamics that aid in model comparisons and also for identifying new studies targeted at better estimation of key parameters.
- **Computation and reporting of prediction variables** – Differences among the models in their prediction variables, including those that overlap across models, is central to a multi-model analysis, but can complicate interpretation of results. For example, the use of different numbers of years in simulations between UBC and NOAA models should be evaluated to ensure that the predictions are robust and sufficiently comparable across

models. The calculation and reporting of P(QET) is especially important because its estimation depends on whether and how stochasticity is represented in different models. Rankings and other measures (e.g., utility) are helpful for interpretation of individual model results but can be misinterpreted, especially when predictions are compared across models. Interpretations of results should always be accompanied by comparing results in native units (observable quantities) to judge the biological significance of differences.

- **Scale of predictions** – Model predictions about the responses to the management alternatives are most robust when used on a relative scale (e.g., changes from the no-action alternative). While model predictions can be viewed on an absolute scale (e.g., numbers of adult individuals expected in nature), they depend on the appropriate benchmarking (calibration and validation) of the model results to observed values from the WVS. Thus, predictions on the absolute scale have higher uncertainty and can be more biased than relative-scale comparisons.
- **Climate change** – The Corps concludes in the Draft EIS that temperature regimes and extreme storm events are likely to change over the next several decades and beyond. The four models did not assess model performance outside the historical range of environmental and biological data for which the models were calibrated. Therefore, robust analyses of the consequences of flow management in relation to such potential changes on Chinook salmon and steelhead are warranted. Climate change will not affect all reaches and tributaries equally, and some local populations may be more at risk than others, both in freshwater and the ocean. Including the likely future temperature and flow regimes in the analysis could shift the ranking and priority of the alternatives. Models that identify such differences in relation to climate trends can inform both short-term and long-term management strategies.
- **Going forward** – The individual models can provide valuable insights into critical uncertainties and how to improve model predictions. This is a rare opportunity when multiple models can be used to improve individual models and their application, and to robustly identify data gaps to inform future data collection. For the future, the ISAB suggests that monitoring and field-based process studies should be continued to improve understanding of fish responses (e.g., survival and movement) to ecological variability within the Willamette River basin. Such studies and long-term data specific to the WVS would provide better inputs and improve validation and calibration for all the models, especially for steelhead. It is likely that the most substantial improvements for the models would come not from modifying the model processes *per se*, but rather by

improving the model inputs and parameters to better represent natural variability and to be more reflective of the WVS.

In summary, the four primary models for spring Chinook salmon and steelhead (and their coupling to the supporting models) are scientifically sound. The strengths and weaknesses of the four models (identified in the individual model reviews) and how they are combined in the multiple model analysis (section II) should be recognized in assessing management alternatives for the Draft EIS. The multiple models provide more confidence than if just a single model was used in the analysis and such multi-model analyses are most effective when the models are implemented in a strategic manner that results in logical and transparent cross-model interpretations. The multi-model approach used by the Corps to date is an excellent approach for assessing alternatives in the WVS EIS process. The ISAB comments on the individual models and on the multiple model approach are designed to further leverage this opportunity of having multiple models, so they further improve the modeling results for informing Chinook salmon and steelhead responses to the EIS flow management alternatives in the Willamette River basin.

ISAB Review of Spring Chinook Salmon and Steelhead Models for Assessing Responses to EIS Alternatives for the Willamette Valley System

I. Introduction

A. Review Request

On April 16, 2021, the Portland District, U.S. Army Corps of Engineers (Corps) [requested](#) scientific review by the Independent Scientific Advisory Board (ISAB) of life-cycle models and other analytical tools used for assessing the response of spring Chinook salmon and winter steelhead listed under the U.S. Endangered Species Act (ESA) to alternative actions considered for the [draft Programmatic Environmental Impact Statement \(DPEIS or Draft EIS\) for the Willamette Valley System](#). On June 21, 2021, the ISAB Administrative Oversight Panel¹ approved the review request and the ISAB officially began to scope its review.

B. Willamette Valley System and EIS Background

The Willamette Valley System (WVS), operated by the Corps, consists of 10 high-head federal dams and reservoirs, 3 run-of-river dams and reservoirs that function as re-regulating projects, and 42 revetments along Willamette River tributaries (Figure 1). The Corps manages storage and release of water from the 13 reservoirs throughout the year to balance and meet the WVS authorized purposes, which include flood control, navigation, hydropower, fish and wildlife, irrigation, recreation, water quality, and municipal and industrial water supply. The Bonneville Power Administration (BPA) is responsible for marketing and transmitting power generated from the eight power-producing projects. The U.S. Bureau of Reclamation (BOR) administers a water marketing program for water stored in the Corps' reservoirs to agricultural users.

As noted above, a key purpose of the WVS is to reduce flood risk, with 27% of the Willamette watershed's runoff regulated by the dams. The elevation change at the 10 high-head dams ranges from 360 to 450 feet from forebay to the average tailwater, and water surface elevations of the reservoirs fluctuate by as much as 160 feet annually, which the Corps acknowledges create significant challenges for fish passage.

¹ The ISAB Administrative Oversight Panel representatives at the time of the approval of the assignment included Aja DeCoteau, Executive Director, Columbia River Inter-Tribal Fish Commission; Kevin Werner, Science Director, NOAA-Northwest Fisheries Science Center; and Richard Devlin, who was then Chair of the Northwest Power and Conservation Council but has since completed his Council appointment.

The Willamette River Basin



US Army Corps of Engineers®
Portland District

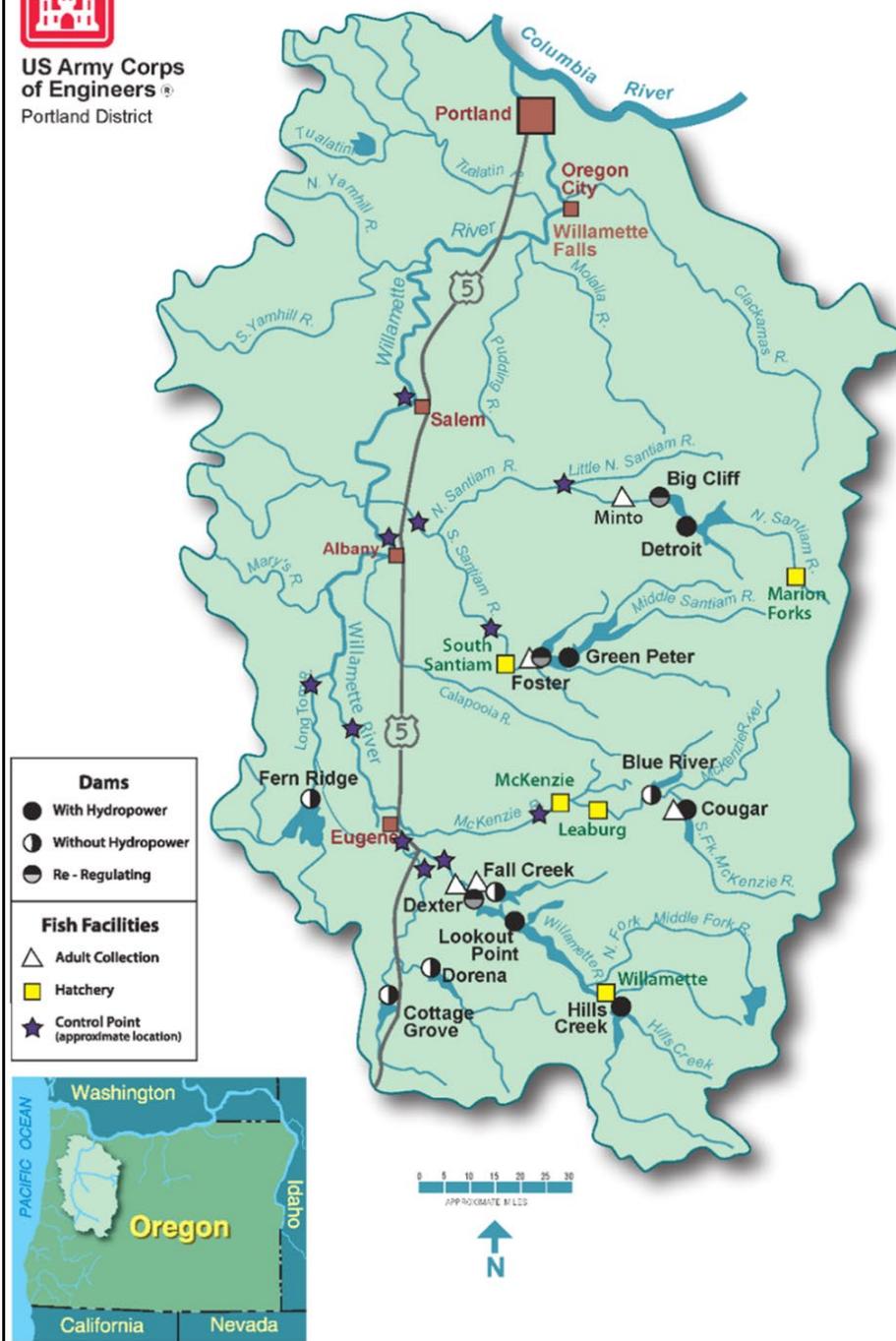


Figure 1. The Willamette River Basin, showing Willamette Valley System projects, including dams and reservoirs, adult fish collection facilities, and fish hatcheries (Source: USACE DPEIS 2022). Also see the Willamette Valley Project Brochure, [USACE 2019](#).

To meet requirements of the U.S. National Environmental Policy Act, the Corps, as the lead federal agency, is preparing the Programmatic Environmental Impact Statement (PEIS)² to assess the environmental effects of the continued operations of the WVS and maintenance needed to meet the authorized purposes. Concurrent to the EIS process, the Corps, Bonneville Power Administration, and Bureau of Reclamation have re-initiated formal consultation under Section 7 of the ESA for the WVS System. The EIS process and analyses provide the basis for a biological assessment as part of an ESA Section 7 consultation process, which will be submitted to the National Marine Fisheries Service (NMFS) and the U.S. Fish and Wildlife Service (USFWS).

A more detailed description of the WVS, its attributes and history, and the EIS process is provided in the [Willamette Valley System Operations and Maintenance: Draft Programmatic EIS, November 25, 2022](#) (hereafter Draft EIS; see the [Corps' webpage](#) for links to the report and appendices). We refer readers to the Draft EIS for additional context.

C. Review Process

The ISAB's review was conducted over the past 16 months while the Corps, its modeling teams, and collaborators conducted analyses and discussions to develop the Draft EIS. The ISAB's review is restricted to providing scientific feedback on the analyses used to inform the Draft EIS and does not include an evaluation of the EIS alternatives or the Corps' decisions on them.

The ISAB reviewed the following primary models used for the Draft EIS:

1. University of British Columbia Integrated Passage Assessment Models (UBC models)
2. NOAA Upper Willamette River Life Cycle Models (NOAA models)
3. ICF Ecosystem Diagnosis and Treatment Models (EDT models)
4. Oregon State University Flow and Fish Survival Models (OSU models)

The Corps used the NOAA, UBC, and EDT models to evaluate the effects of EIS alternatives on spring Chinook and winter steelhead populations in the North Santiam, South Santiam, McKenzie, and Middle Fork Willamette river basins. The OSU models focused on within-year effects of flow management.

In addition to these primary models, the sources of key inputs for the models were considered for context. Specifically, we examined how the Corps' used the updated Fish Benefits Workbook (FBW) to provide estimates of downstream passage survival for Spring Chinook salmon and steelhead for the primary models. The ISAB previously reviewed the FBW in 2014 ([ISAB 2014-3](#)). The Corps also uses the HEC-ResSim hydrology model ([DPEIS Appendix B: Hydrologic Processes](#)

² The EIS is "programmatic" because it is at the program level covering the full system rather than the site-specific project level. Site-specific NEPA analyses occur when projects are proposed. See the DPEIS for details.

[Technical Information](#)) to estimate hydrological information and USGS CE-QUAL-W2 model ([DPEIS Appendix D: Water Temperature and Dissolved Gas Methodology](#)) to estimate water temperatures to support the primary models in the EIS analyses. We considered them for context but did not specifically review them. The four primary models and other analytical tools reviewed by the ISAB are documented in *DPEIS's Chapter 3: Affected Environment and Environmental Consequences*, as well as [Appendix E: Fish and Aquatic Habitat](#).

To review the models, the ISAB developed a detailed set of questions, which integrate and expand upon questions included in the [Corps 2021 review request memorandum](#), follow-up questions provided by the Corps, the ISAB's 2014 review of NOAA's Willamette model, and internal ISAB discussions. The full set of questions and sub-questions are listed in the Summary Answers subsection below and in Appendix 1 of this report. The questions are organized under the following topics:

A. Individual models

1. Model structure and development
2. Key assumptions and limitations
3. Model validation and verification
4. Data gaps
5. Model output
6. Model improvement

B. Use of multiple models

1. Evaluation of management strategies
2. Model coupling
3. Model components
4. Information representation
5. Model comparisons
6. Using the models after the completion of the EIS
7. Communication to managers and stakeholders

The ISAB's review of the models began in November 2021, at which time the models were in different states of completeness. This ISAB report was completed after the Draft EIS was released in November 2022. Over the past 16 months, the ISAB met with each modeling team, reviewed materials as they were provided to us, shared clarifying questions with the modeling teams and the Corps, and reviewed the Corps' and modeling teams' responses. The staggered review included numerous interim milestones: (1) initial written materials for the OSU models and FBW were submitted and a presentation was provided in November 2021 (supplemental material was also provided in June 2022); (2) introductory presentations on the three other

models were provided in February 2022; and (3) written materials were submitted for the NOAA and EDT models in June 2022 and the UBC models in October 2022. The meetings and response discussions between the ISAB and modeling teams informed our review, answered our questions pertaining to the individual models, and were intended to assist the Corps and the modeling teams in refining their analyses, improving their scientific approaches, and coordinating their modeling efforts and EIS analyses. An important outcome of these iterative discussions was that the ISAB was able to share feedback on individual models that could be considered as the modeling teams developed and documented their models.

Although the ISAB was able to work with this sequence of review materials, the lack of synchrony delayed and complicated the ISAB's understanding of how the Corps used the different models individually and collectively to inform the EIS process. At the same time, the ISAB is sympathetic to the Corps' challenge of working with four different modeling teams and developing the Draft EIS. In spite of the varied timing of review document availability, the ISAB attempted to provide rigorous reviews with consistent attention to technical details for each model. Ultimately, we received the information needed to complete a thorough scientific review and greatly appreciate the Corps' coordination and communication. The four modeling teams' provided essential documentation and constructive responses to the ISAB's inquiries throughout this review process.

To produce a timely review in light of the sequential availability of model reports with the final combination of results occurring last, the ISAB reviewed each of the models separately as if they were the only model used in the analysis. We then reviewed how the model results were being combined for the WVS EIS in the Corps' multi-model approach. Therefore, some of the ISAB's comments on the individual models become moot in terms of their application for assessing the EIS alternatives because they were addressed by how their results were ultimately used as part of the final multi-model analysis. Reviewing the individual models as standalone applications is useful because each review is complete, use of the models in the multi-model approach can change, and the models may be used as stand-alone applications for the WVS or in other similar situations in the future.

The ecological and river management issues involved in the Corps' management of the WVS reflect many of the challenges addressed in the Fish and Wildlife Program of the Northwest Power and Conservation Council. This ISAB review builds upon previous ISAB reviews of analytical approaches for the Willamette River basin: *Review of NOAA Fisheries' Viable Salmonid Population (VSP) Modeling of Willamette River Spring Chinook Populations* ([ISAB 2014-4](#)) and *Review of the Fish Benefits Workbook for the U.S. Army Corps of Engineers Willamette Valley Project* ([ISAB 2014-3](#)). Other past science reviews are also informative, including the ISAB's *Review of NOAA Fisheries' Interior Columbia Basin Life-Cycle Modeling Draft*

*Report ([ISAB 2017-1](#)) and the *Independent Scientific Review Panel’s Review of the Research, Monitoring, and Evaluation Plan and Proposals for the Willamette Valley Project* ([ISRP 2011-26](#)).* These topics are relevant for other ISAB and ISRP assignments as well. The Corps’ analyses and efforts address reintroduction of anadromous fish to areas blocked by high-head federal dams, a major management challenge that is being discussed elsewhere in the Columbia River basin (e.g., above Chief Joseph and Grand Coulee dams ([ISAB 2022-2](#))).

D. Report Organization

Section II of this report provides a review of the multiple models used by the Corps for the WVS Draft EIS and is intended for agency policy analysts and the modeling teams. The final section includes more specific, technical comments on the individual models and is intended to provide information to assist the EIS authors and the modeling teams as they improve the models and their documentation in the future.

II. ISAB Comments on the Use of Multiple Models to Assess ESA-listed Fish Responses to Willamette Valley System EIS Alternatives

A. Background

For the purposes of the EIS and the ongoing ESA Section 7 consultation, the Corps are using an evaluation process to analyze the cumulative effects of reservoir operation alternatives on listed species that combines the results from multiple models. A multiple-model approach can effectively quantify the level of uncertainty of predicted responses to alternative operational scenarios. While modelers often perform sensitivity analysis that varies individual parameters of a model, it is much more difficult to assess the sensitivity of model predictions to alternative assumptions and formulations (i.e., structural uncertainty) (Refsgaard et al. 2006; Parker 2013). In concept, each model represents a different (but plausible) view of the real system. Judging which model is best is often difficult because each is based on defensible assumptions and may perform similarly when compared to field data as part of calibration and validation.

When predicted responses of multiple models agree, uncertainty is reduced, increasing confidence in the model results. Disagreements among models are also useful because they show a range of possible responses that are plausible. Estimation of uncertainty with multiple models (whether low or high) is more accurate than if a single model was used without awareness of what alternative models would have predicted. A single model allows for parameter uncertainty but not structural uncertainty, while multiple models allow for quantifying both. This more accurate understanding of uncertainty in predictions helps determine: (1) whether model responses differ among alternative management actions and (2) the likelihood the response from a specific management action will be realized in nature. If model structure uncertainty is large, then the wide range of predictions for each management alternative can mask differences among them. With one model, a user can mistakenly infer that the responses of alternatives are different. Similarly, a wide range of responses for a selected alternative implies that the expected response in nature is highly uncertain based on our present knowledge, and thus there is greater chance that the predicted response will not occur. Therefore, multiple models inform robust decisions, whether they agree or not. An important aspect of multi-model analyses is to also understand the underlying reasons why the multiple models agree or not.

Multiple model approaches are widely used in many fields. Well-known applications include hurricane tracking (e.g., Hamill et al. 2012; Lin et al. 2020) and global circulation modeling for climate change (e.g., Fordham et al. 2012; Parker 2013; Hersbach et al. 2015). There are also

examples of using these approaches for ecological forecasting (Bonan and Doney 2018) and evaluation of fish population responses to inform management (Jardim et al. 2021). Typically, classical statistics evaluate alternative models (often systematically increasing their complexity by the number and functional representation of variables) by using metrics to select the “best” model (e.g., Akaike Information Criterion, AIC). In contrast, the goal of multiple model analyses like the above examples (and the Willamette application) is to capture the uncertainty from the models rather than select the best model. Another major difference is that alternative models can be rapidly fit to the data with the classical statistical approach, while simulation models, such as life cycle models, require a large effort to develop, run, and analyze.

The success and usefulness of multiple models in reducing uncertainty depend on how the models are selected and implemented and how results are interpreted in a specific situation. Key aspects of a multiple model analysis are: why the different models were selected; how much information they share; the operating constraints for calibration, validation, and scenario analyses; and how the predictions are combined across models to inform decisions. The extremes range from zero exchange of information among the models and modelers to the other extreme where all common information among the models is shared and used the same way (e.g., Rose et al. 1991a,b). Most examples are somewhere on the continuum between the two extremes.

Importantly, there is no single correct or incorrect way to do a multiple model analysis. What determines effectiveness and success is how the analyses are conducted and interpreted. For example, a common mistake is to use multiple models and then simply treat their predictions as completely independent. This can lead to overconfidence when predictions agree because the appropriate confidence level is less because the models are not independent. Thus, treating agreement from multiple models as if they were independent inflates the confidence level. This can lead to overly optimistic expectations because it appears as if multiple models predict the same response. In some cases, such misinterpretation cannot only lead to higher risks of a given decision not being realized in nature, but it can also lead to selection of a poorer alternative. At the other extreme is forcing agreement among models due to the models selected being essentially versions of the same model or having them share too much information. In this case, their predictions essentially provide the same confidence as if one model was used. Agreement among the models in this case also leads to over-confidence in model predictions. Multiple models, if carefully and thoughtfully implemented and interpreted, are a powerful approach to quantifying uncertainty and providing more accurate information to inform decisions.

Detailed reviews of each of the individual models are provided in later sections of this report. Those reviews were conducted as if the specific model was being considered independently of

the other models. In this section, we focus on how these models were used together in the EIS process or could be used in the future. The Corps plans to update the Draft EIS after the public comment period and also to use the multiple models in future analyses and management decisions. Thus, the ISAB provides the following suggestions and considerations for the Corps as they continue developing and applying the multiple models.

B. Multiple Models used by the Corps for the Willamette Draft EIS

The multiple model analysis consisted of four primary models and several supporting models, all being used in a coordinated way to assess management effects on upper Willamette River Chinook and steelhead. All these models are used to inform management decisions about selecting from alternative scenarios of operations. Three of the primary models (NOAA, UBC, EDT) generate prediction variables of population responses (Table in Figure 3) that are part of the VSP (Viable Salmonid Populations; McElhany et al. 2000) set of variables (adult abundance, productivity, spatial structure, diversity). The fourth primary model (OSU) also generates prediction variables used to inform the decision-making, but the variables are not within the VSP set. Each of these models relies, to various degrees, on the output of other models (termed supporting models). The primary models use the outputs from the supporting models as their inputs including: (a) Fish Benefits Workbook (FBW) for juvenile fish dam passage survival and timing estimates, (b) the USACE (Corps) Hydrologic Engineering Center ResSim model (ResSim-Flows) for reservoir operations affecting river flows, and (c) the Portland State University Water Quality Research Group CE-QUAL-W2 hydrodynamic and water quality model for stream temperatures.

Figure 2 shows the exchange of information from the supporting models to each of the four primary models. Operationally, the supporting models were run and then the four modeling teams operated mostly independently. The modeling teams were tasked with the same questions and scenarios and were provided with the same outputs from the supporting models. A key commonality among the EDT and OSU models was the required use of three specific years (2011 - wet, 2015 - dry, and 2016 - normal) in their analyses. Each team documented their modeling, decisions, and results in reports and sent their predictions for the populations (subbasins) and alternatives to the Corps.

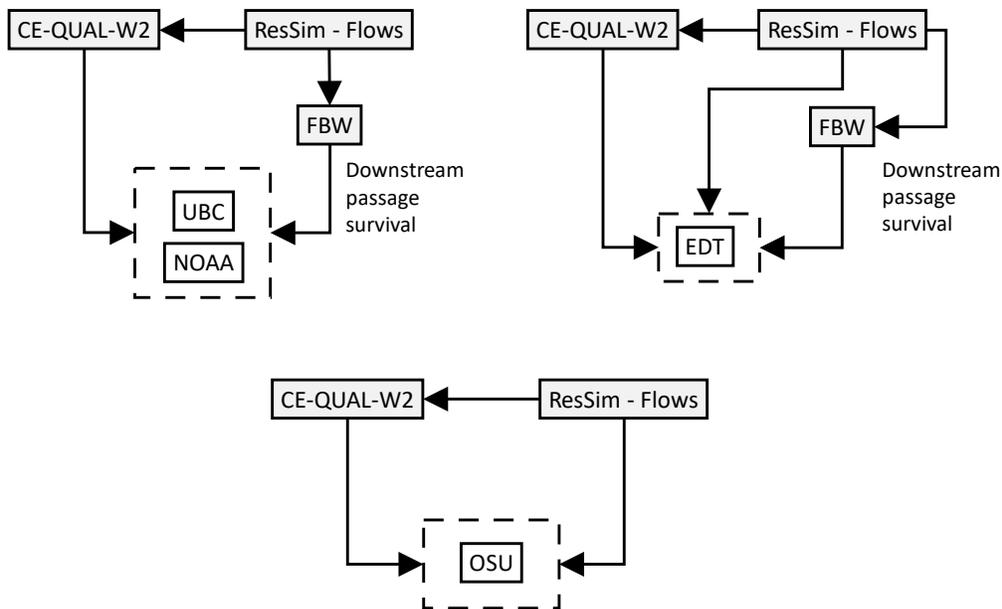


Figure 2. Exchange of information from the supporting models (grey boxes) to each of the primary models (white boxes) in the Corps analysis for the WVS EIS.

The predictions from the primary models were collated by the Corps and then combined to evaluate the population responses to the alternative scenarios (Figure 3). The overall synthesis involves combining results across years of simulation, across prediction variables (metrics), and across subbasins for a model. The final results are then summarized and reported for each alternative. The UBC and NOAA models generated comparable predictions of adult abundance, productivity, and extinction risk, while the EDT model also generated predictions of abundance and productivity and additionally, predicted diversity (Table 1 in Figure 3).

The combining of results starts with a table for each population and alternative. Each prediction variable from NOAA and UBC models is then rated as “fails” or “meets” based on a defined acceptable threshold for each prediction variable that was established by the Corps for this project (Table 2 in Figure 3). The thresholds for the NOAA and UBC models were (1) abundance was greater than the sustainable or recovery levels; (2) productivity (R/S) was greater than 1 (0.9 for NOAA because averaged over 100 years), and (3) extinction risk was less than or equal to 5%. For EDT, the table shows the rank (1, 2, or 3) of the three separate-year simulations (wet, normal, dry). Thus, model results are presented as a table of “fails-meets” across the three prediction variables for NOAA and UBC models, and as a 1, 2, or 3 as the ranks for the wet, dry, and normal year results for the EDT model. The EDT rankings indicate how robust the different prediction variables were for that specific alternative to variation in hydrology. For example, if the ranking (1 = best to 3 = worst) was consistently in the order of wet being the best and dry being the worst across the three EDT prediction variables, then the alternative is robust because all prediction variables behave the same way in the alternative in response to

different hydrological year types. Under such high consistency, one does not have to trade-off the performance of an alternative among the prediction variables. For example, complete consistency to hydrology avoids the situation of an alternative being ranked best for abundance but worse for diversity having to be compared to an alternative that is intermediate for both. Such situations then require a weighting scheme to evaluate the trade-offs of which prediction variable is more important under different hydrological year types: wet, normal, and dry.

Each population and alternative combination was then rated (Table 3 in Figure 3) using an effects scale (none/negligible, minor, moderate, major) based on the meets/fails tallies from both NOAA and UBC models, rankings from EDT, and the number of persistent populations (basinwide). These ratings of none to major are from the perspective of the magnitude of adverse effects that an alternative would have if implemented, and the Corps defined the ratings for the Draft EIS. The McKenzie population is considered a legacy population and high value is placed on its persistence because it is considered to show the full expression of juvenile life history migratory strategies (page 3-667 of DPEIS). To illustrate the criteria for the ratings, a rating of Minor is when: *Sub-basin extinction risks are between 1% and 5%, maximum productivities are greater than 1; equilibrium abundances are above the levels needed for recovery, and 3 of 4 populations would persist with one being the legacy population.* The last criterion is based on looking across populations for each alternative, and it is a count of populations that show persistence and also considers whether the legacy population was persistent. Because this is based on a count of populations throughout basin, a single basin-scale value is used for all populations for an alternative. A basin-scale value is used for an individual population because it reflects the extinction risk of all populations, including the population of interest and the legacy population, for that alternative. When an alternative meets the criteria (i.e., high viability), this indicates small (none or minor) adverse effects; as an alternative fails more criteria, the rating shifts to moderate and then to major adverse effects.

The Corps used professional judgment in the interpretation based on viewing which model predicted the fails and for what variable and alternative. For example, the South Santiam River under Alternative 1 (see Table 3.8-26 in DPEIS) and the McKenzie River under Alternative 1 (Table 3.8-28) generated identical meet/fails for the NOAA and UBC models and only differed in their rankings of wet, normal, and dry years by EDT for diversity. Adverse impacts were rated as MINOR for the South Santiam River and MAJOR for the McKenzie River. The logic of the ratings was clearly explained for each population-alternative combination in the EIS. In general, such interpretation of multiple models results is appropriate, provided they are well supported and clearly documented.

A second path for combining model results was also used in the overall synthesis to generate a decision matrix (right-hand side of Figure 3). The decision matrix table has a column for each

alternative and comes directly from the model predictions (not the “meets/fails”). The prediction variables or metrics for each alternative, relative to the results for the No Action Alternative (NAA), are reported as: (a) change in the number of populations reaching replacement (productivity), (b) change in the number of populations deemed persistent (abundance), (c) whether the risk of extinction was reduced or increased, including the legacy population, and (d) rank of downstream survival among the alternatives. Metrics (a), (b), and (c) were used from the UBC model (the NOAA model has not been used in the EIS to date) and (d) was obtained from OSU model flow-survival results that ranked performance (survival below dams as affected by flow and temperature, see page 5-6 of the EIS) across the alternatives (i.e., 1 to 7). In this ranking scheme, higher rank indicates that alternative shows better ecological performance.

Our review of the use of multiple models focuses on Chinook salmon and the prediction metrics related to the models being reviewed. A similar (but not identical) approach was used for steelhead, which is not detailed here. Water quality (temperature, TDG, turbidity, and mercury) was also evaluated, but only temperature was derived from a model considered here (CE-QUAL-W2 supporting model), which the ISAB did not review; and multiple models were not used.

The following section focuses on strategies and considerations for implementing and interpreting the multiple primary to maximize their usefulness for informing decisions. The Corps has considered many of these suggestions to various degrees, but these considerations either would benefit from further elaboration or their documentation has not been synthesized.

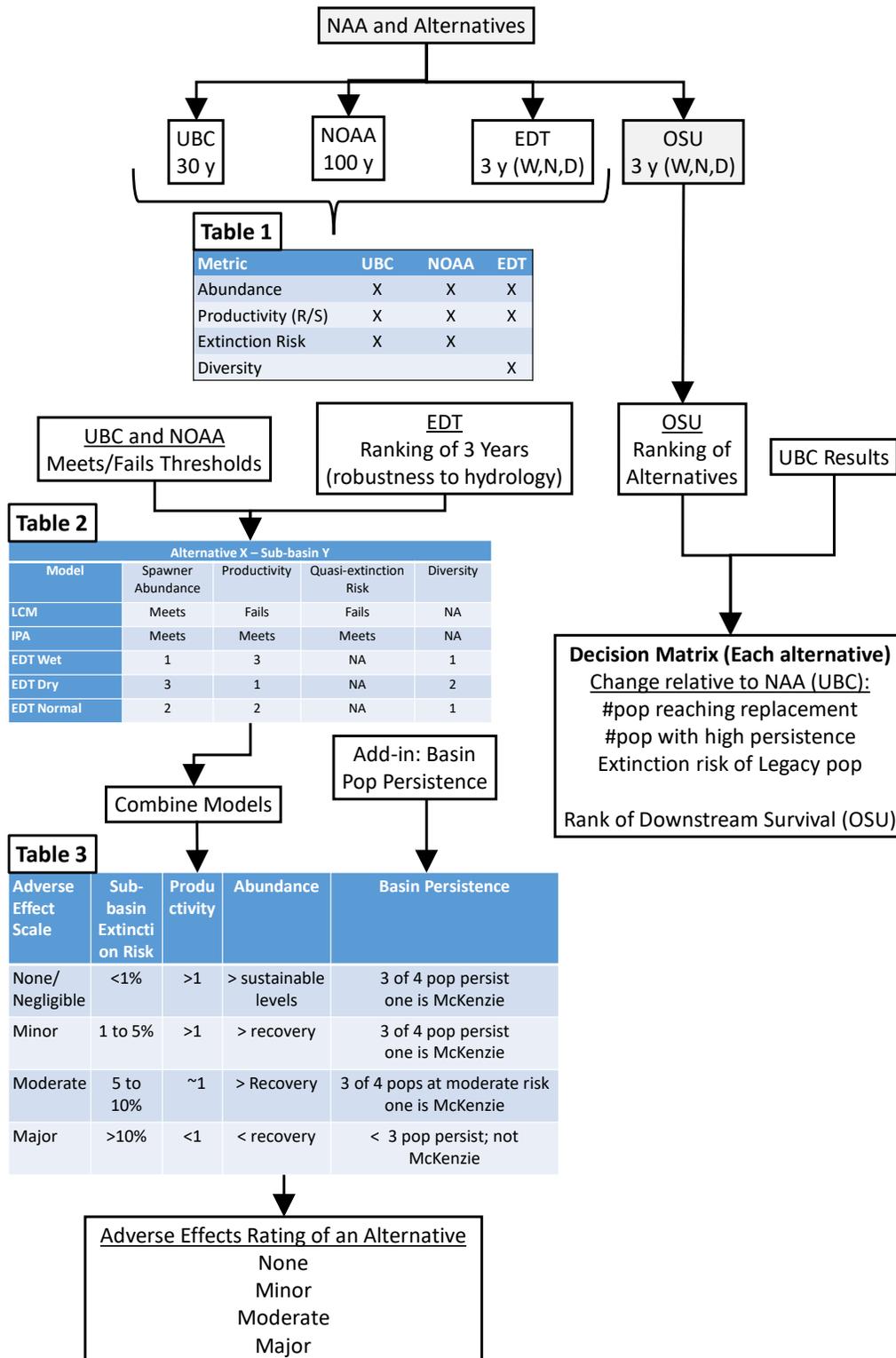


Figure 3. Flowchart showing the steps in the overall synthesis of the multiple models.

C. ISAB Comments

1. Model selection and differences

A critical step in multiple model analyses is documenting why the specific models were selected from all possible models, and in this case, why a new model (UBC) was developed. Knowing why the models were selected, their similarities and differences, and their history of development and applications is useful for interpreting the model results. Ideally, available models are reviewed, and their assumptions and structure tabulated in sufficient detail, so that the selection is strategic and clearly documented. The models should overlap in their inputs and predictions to some degree (so they can use some shared inputs and address at least some of the same questions), while also being different but plausible representations of the same system in nature. An important aspect of this documentation is also a description of the lineage of the models (and sometimes the lineage of the training and collaborations of the modelers). Models often evolve over time and generate new versions that then may appear as new models or as models influenced by earlier models (e.g., UBC model resembles the NOAA model). Also, it is often difficult to separate the model from the modelers (see Thompson 2022). Application of simulation models involve many decisions and thus shared training (e.g., same advisor, research group) and collaborations of the modelers can lead to more similar models than if they were developed independently.

Suggestion 1: Prepare a document that is a detailed comparison of the assumptions, processes included, formulations (e.g., mortality dependent on size), history of development, and summary of the reasons and logic for selecting the models. An example of a model selection document related to large-scale ecosystem restoration analogous to the Willamette situation can be found in Rose and Sable (2013). This document was followed by subsequent documents on implementation and design of the multiple model analysis and the interpretation of their results (see Lewis et al. 2021).

Suggestion 2: Collate results from previous applications of the selected models to qualitatively evaluate past model performance and the types of questions addressed. This information can provide additional insights into the strengths and weaknesses of each of the models, and therefore help in the interpretation of their results when used with the multiple model analysis. For example, in the EIS there are criteria based on absolute values of extinction risk (e.g., 1%, 5%, and 10% cutoffs). Yet, these models can differ greatly in their ability to estimate absolute extinction risk, as discussed in the model reviews below. Thus, more information on the capabilities of the models to predict extinction would aid in the interpretation.

2. Configuration and shared information

After the models are selected and their differences and capabilities are identified, the configuration step involves moving to the specific application, determining how the models are set-up, providing shared inputs, defining scenarios, and identifying requirements of predictions to facilitate cross-model comparisons. As part of shared information, all primary modeling teams were aware of the overarching question that the modeling should address: How do the designated prediction variables related to Chinook population dynamics differ among the alternatives for the populations of interest?

An important aspect of configuration is what information is shared among the models. In the Willamette EIS analysis to date, the supporting models provided temperature, flows, and downstream survival information and were the source for differences between alternative management actions. The modelers were required to generate predictions for the pre-defined alternatives defined by the Corps. However, they were free to use the provided outputs from the supporting models, and to design the specifics of their simulations with their models, as they deemed appropriate. The prediction variables from the primary models were mapped to each other and to the variables used in VSP (Figure 3). Note that to help in the interpretation, the prediction variables comprising VSP were used separately in the EIS and not collapsed into a single measure as is sometimes conducted with other analyses based on the VSP.

The models sometimes generated different versions of the same prediction variable. For example, adult abundance was computed as: (a) a mean (over runs) of the geometric average over years 1 to 100 in each run for the NOAA model, (b) the median (over runs) of the geometric average for years 16 to 30 in each run for the UBC model, and (c) the equilibrium value for the EDT model. The modelers could decide the number of years, how to populate the run with historical conditions from years covered by the supporting models, and how to report the final predictions. While some models reported their own output summary (e.g., rankings), output files with the prediction variables in native units were made available to the Corps. The availability of native unit results is important because use of ranks and utility can create answers that rely on exaggerated differences among the scenarios. Indeed, the purpose of utility functions is to highlight differences by rescaling original units into zero-to-one regardless of how different the predictions were across alternatives (DeWeber and Peterson 2020). Such rescaling of model outputs can be useful for comparisons within a single model but hinder inter-model comparisons.

Use of shared information raises the possibility that the outputs from one of the shared sources (supporting models) used as inputs to the primary models dominate the predicted differences among the alternatives in two or more of the primary models. For example, how much do the

results of the FBW control the predictions of the NOAA, UBC, and EDT models? In situations where results are dominated by a single supporting model, the differences among the primary models (i.e., structural uncertainty) play a diminished role and the major uncertainty shifts to how well the supporting (single) model predicted differences among alternatives. Typically, when appropriate input data are available, models of hydrodynamics and hydraulics generally have lower uncertainty than ecological and fish models. However, the ability of the supporting model to distinguish among the alternatives becomes critical when a supporting model drives the differences among the primary models.

Suggestion 3: Document how each of the models was configured and run using a format for the document that facilitates comparisons across models. Much of this documentation could be obtained from information available in the EIS, from presentations, and from the Corps as they continue the multiple modeling.

Suggestion 4: Provide clear graphs and/or tables of the outputs of the supporting models, and then indicate how they were processed and used in each of the primary models. For example, the NOAA model did not use the temperature provided as direct input but, rather, applied the temperature outputs to compute changes in temperature and applied those changes to the temperatures used in their model. Knowing how the adjusted temperatures in the NOAA model compared to the temperatures used in the other models is important. A presentation of the outputs of the supporting models and the actual values of those used as inputs in the primary models would be useful for understanding model prediction similarities and differences.

Suggestion 5: Show how differences in predictions among the alternatives from the primary models relate to the differences in the model structures or decisions in how they were implemented. This can be done by detailed examination of *why* the models generated similar or different predictions or by new simulation experiments that test for the importance of specific supporting models. For example, to evaluate the importance of the FBW, the primary models could be run with only FBW results used, holding other inputs unchanged from their reference values to contrast alternatives. If the predictions for alternatives are similar between when only FBW changes are imposed compared to when inputs from all supporting models are changed, then it can be established that FBW results, rather than structure or parameterization of primary models, determine the predicted differences among alternatives.

3. Independence of predictions

The models' predictions are not completely independent from each other because they were developed with knowledge of the other models and they shared some common information (same field data, supporting models). However, the multiple models provide more confidence than if just one of the models was used in the analysis. For example, the use of two models for

a prediction variable (e.g., abundance by NOAA and UBC models) has the confidence level somewhere between one and two independent models. Two models with similar structures and that share critical information are likely closer to one independent prediction, whereas two very different models with minimal shared information may be closer to two independent predictions. The degree of independence can be judged based on the structural differences in the models and the constraints imposed on the configuration of the models. The goal of multiple modeling is to better characterize uncertainty and to avoid using the multiple models in a way that results in under or over confidence.

Suggestion 6: While determining a precise value for how the multiple models correspond to the number of independent predictions is not possible, one can qualitatively evaluate whether some models are structurally more similar to others, the degree to which common inputs influence results, and diagnostics on why models predicted differences among alternatives. Therefore, we recommend that a study be conducted to assess the extent of model independence by: (1) comparing process representations, (2) evaluating the past performance of the models when applied to similar questions, and/or (3) performing new analyses with the models that are specifically designed to assess how their structural differences affect predictions.

4. Combining results

Predictions from the primary models can be combined in several ways so that the collective predictions can better inform decisions. The Draft EIS used multiple models for the impacts ratings and only used one model (UBC model) for the decision matrix. The results of all primary models were considered but not all information was documented in the Draft EIS. The primary models generated predictions for the same seven management (operational) alternatives but differed in the number of years simulated and how the three historical years from the supporting models were represented within their simulated historical time period. They also used the output from the supporting models for alternatives, but not all models used all of the outputs provided from supporting models and some primary models used the same outputs differently. These differences are a valuable part of multiple model analyses and are the strength of multiple modeling if the results are combined in a structured and transparent way.

For the results to date for the Willamette EIS analysis, and for most applications of multiple models, the options for combining model results are: (1) collate, (2) integrate, and (3) *a priori* select the best. Collating reports the predictions of all of the primary models in the same format and then interprets the results, keeping the predictions of each model separate. Integrating merges and combines the predictions before the interpretation. One can simply average the results using percent changes from each model or use the maximum or minimum values among

models for each prediction variable. Averaging as part of integration commonly uses a weighting scheme that reflects the different level of confidence in each of the models. If a simple average of prediction variables is used, this means that all models are considered equally informative (a very strong and unlikely assumption).

The third option for combining results of selecting the best model is based on the assumptions or performance (past or in this application) of the primary models during calibration and validation. Some prediction variables may be identified as being stronger in some models than in others. In this scheme, the predictions from the best model only are used for each prediction variable that has two or more models. A key aspect of the “select the best model” approach is that the determination of the “best model” is done *a priori* based on the assumptions and performance (calibration, validation) of the models for past and the current applications. Considering the predictions of alternatives that will be used for the decision-making should be avoided in selecting the “best” model. The selection process does not identify the better model but rather identifies the better predictor for each prediction variable. Other versions of combining results and applications can use a mixture of these approaches.

The Corps in the EIS used a hybrid version of these options. They started with a collation and used a mix of integration and selecting the best to assign the ratings. The logic presented with assigning ratings to each population-alternative explained how the “fail-meets” were given different weightings (although not specified numerically) depending on which model and which prediction variable. The ISAB selected 10 combinations of population-alternative presented in the EIS for Chinook to illustrate ways of exploring how the models were combined in the EIS (as illustrated in Figures 4 and 5, generated by the ISAB). In practice, these comparisons would be labeled so population and alternative specifics were known and factored into the interpretation. Our purpose in generating these figures was to explore and illustrate graphical presentations of the results and should not be used to interpret the performance of the alternatives on populations presented in the Draft EIS.

Figures 4 and 5 demonstrate a graphical approach to communicate how the NOAA and UBC model results were used to determine the impact ratings. Figure 4 illustrates the fails/meets results for six prediction variables (abundance, productivity, and risk, for each of two models), color coded by whether the rating was major (red) or minor (blue). If all models and metrics are considered equal (idealized situation that almost never occurs in practice), one would expect blue lines to tend towards the “meet” (top of each plot) and red lines to concentrate near the “fails” (bottom of each plot). Using our 10 combinations, there is much variation in the pattern of the lines across graphs (alternatives), such as the blue lines (minor effect) not always at the top (i.e., a mix of “meets” and “fails”) and the red lines (major effect) not always at the bottom (i.e., not always “fails”). Thus, the adverse effects ratings were not a simple tally of the “meets”

and “fails.” Rather, the complicated patterns show that the interpretation by the Corps relied on which metrics, models, and populations were generating the “meets” and “fails” to determine the adverse effects ratings. Figure 5 re-organizes the same information as in Figure 4 to illustrate the influence of the NOAA model versus UBC model in the ratings. The fewer and flatter lines indicate that the two models generated similar metrics. The second and fourth panels illustrate the extremes: fails for both models for all metrics (major effect) and meets for both models for all metrics (minor effect).

Figure 6 presents a graphical summary of EDT results about robustness to hydrology for the same 10 population-alternative combinations used in Figures 4 and 5. Lines in each plot show the rankings across wet, dry, and normal for each metric predicted by EDT. Red was rated major adverse effect and blue rated as minor adverse effect; shading in each are abundance, risk, and diversity. If only EDT results are used and all metrics are treated equally, then blue (minor effects) should have few flat lines because they overlay and receive the same rankings across metrics (i.e., very robust to hydrology). Major adverse effects (red lines) would show all three lines (no overlay) and they would crisscross across metrics. The lack of simple patterns indicates that the ratings that used all of the primary models were not simply based solely on the robustness to hydrology (i.e., the EDT model rankings alone) and were, therefore, greatly influenced by the NOAA and UBC model results.

Suggestion 7: All model predictions should be provided in native units for all years in all runs. We also encourage the Corps to report intermediate variables in native units because they reveal why the predictions differed among alternatives. Intermediate variables typically relate to each life stage, focus on rates (e.g., mortality), and examine spatially specific aspects of the results. Examples of such intermediate variables are abundances entering each life stage for each year, partitioning of mortality among sources (e.g., dam passage, ocean) represented within the model, and deconstruction of the results spatially within populations and among populations. Other reporting of predictions by the modelers is encouraged because they know the details and strengths of their respective models.

Suggestion 8: Check the ecological relevance whenever using ranked or other re-scaled results. Ranks change differences (among populations, among alternatives) into a uniform scale without showing the magnitude of the differences in the metric or ranking variable. For example, predictions of 5, 10, and 11 gets ranks 1, 2, and 3, which is the same rankings as predictions of 5, 6, and 11 and 5, 6, and 7. Reporting the results in native units, as well as rescaled, will avoid misinterpreting re-scaled model results.

Suggestion 9: Because all results are now available (they were not for the EIS), the Corps has an opportunity to further strategically refine the overall synthesis of the model results. The Corps’

decision matrix reported in the Draft EIS used the UBC model and not the NOAA model. A relatively simple approach would be to follow the NOAA and UBC models, each paired with EDT and OSU models, through to entire analysis (i.e., the collate scheme seems appropriate for the initial integration). Tables provided in the Draft EIS for one model (UBC model) could be prepared for all of the primary models and each followed through the decision-making as if they were the only model. Thus, the sensitivity of the evaluation of alternatives to which model was used could be assessed. Alternative schemes could also be considered in the collation, such as using the range of values (minimum and maximum) for each prediction variable to see if that would affect decisions.

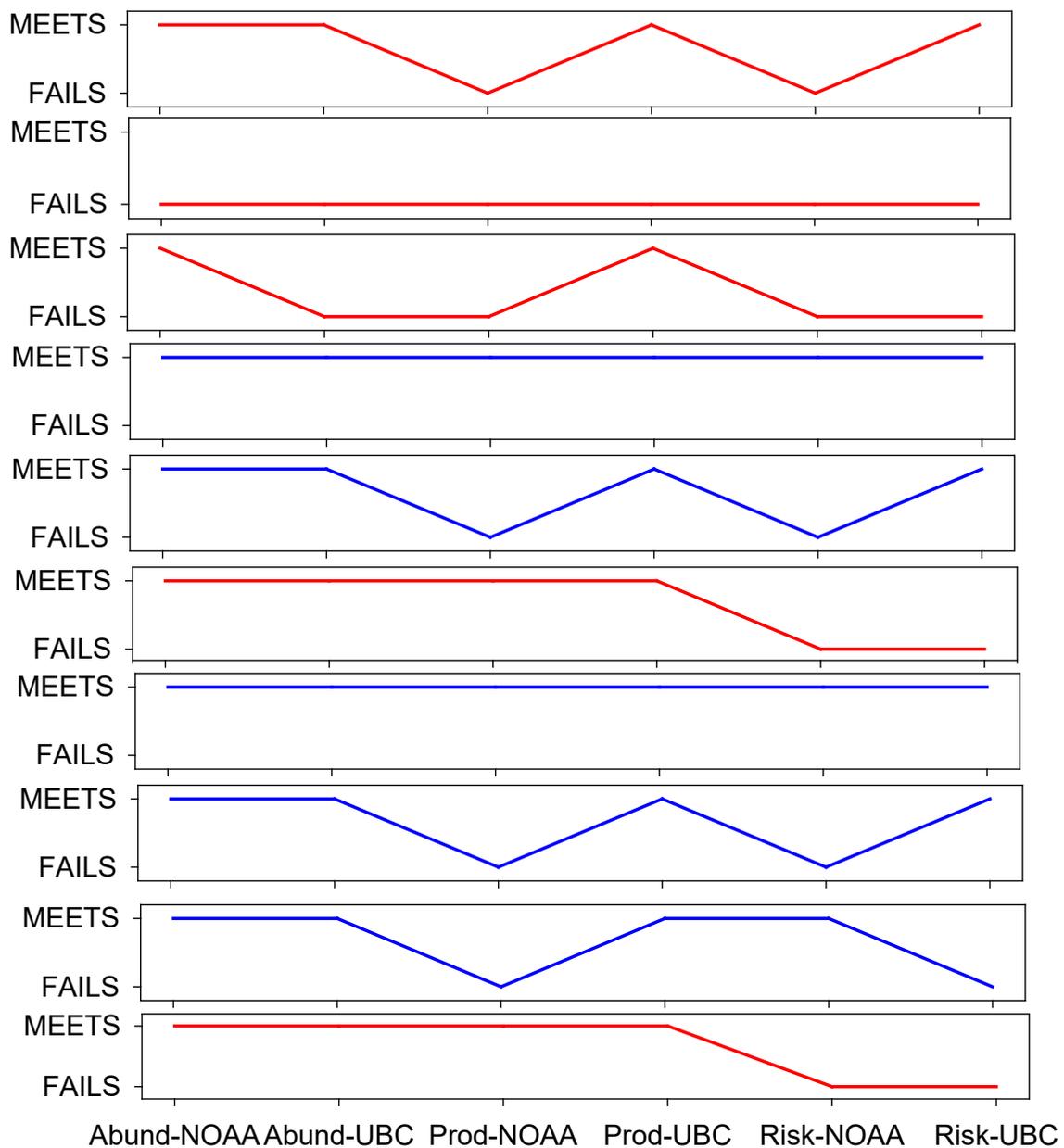


Figure 4. Graphical display of the Meets/Fails tables for the three metrics (Abundance, Productivity, and Risk) for the NOAA and the UBC models, and the resulting adverse effects rating of Major (red) or Minor (blue) impacts for ten selected population-alternative combinations reported in the Draft EIS.

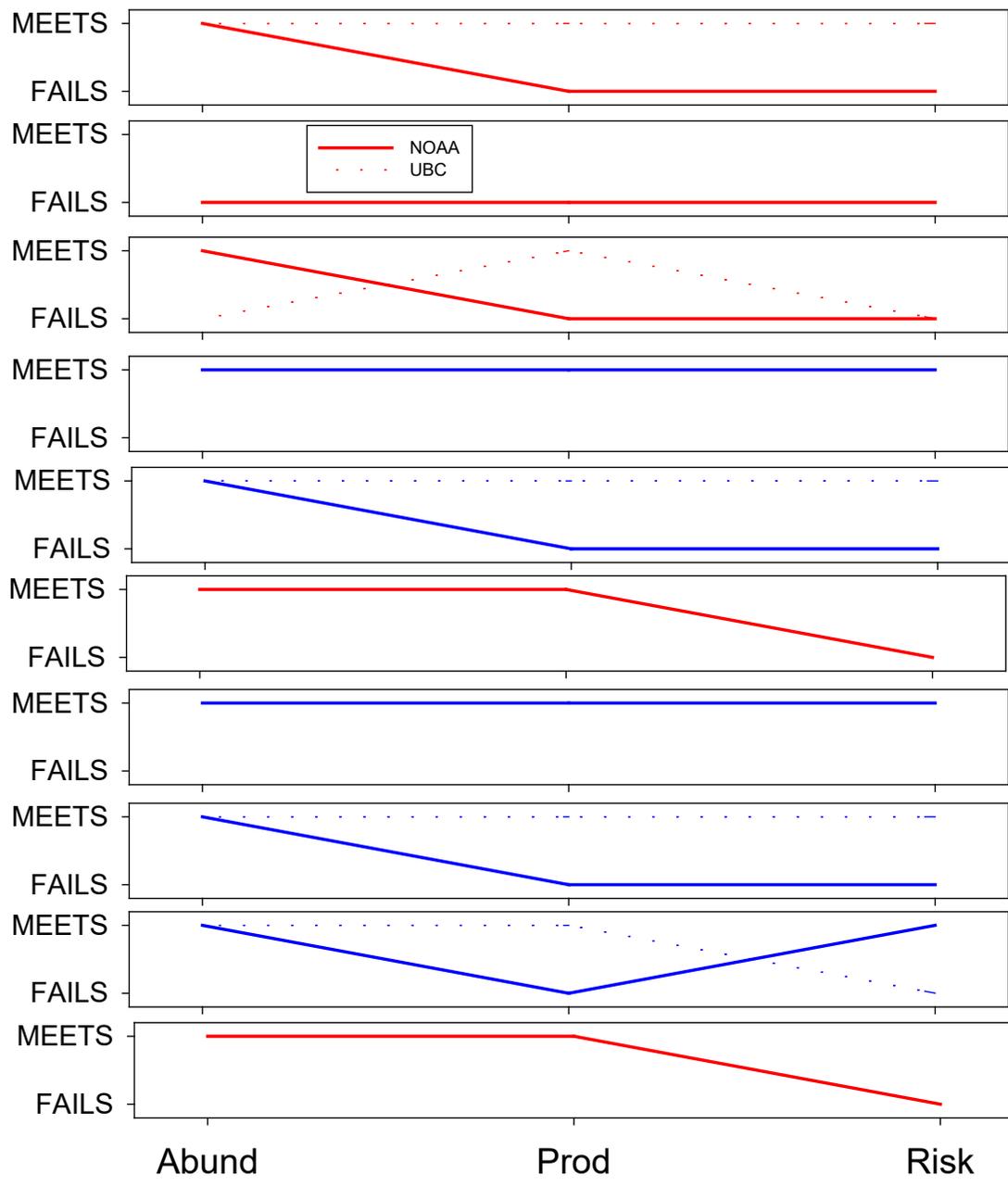


Figure 5. Graphical display of the Meets/Fails for the same ten selected population-alternative combinations shown in Figure 4 showing the three metrics (Abundance, Productivity, and Risk) by model (NOAA model is solid; UBC model is dashed). The adverse effects rating is the color of the lines with Major as red and Minor as blue.

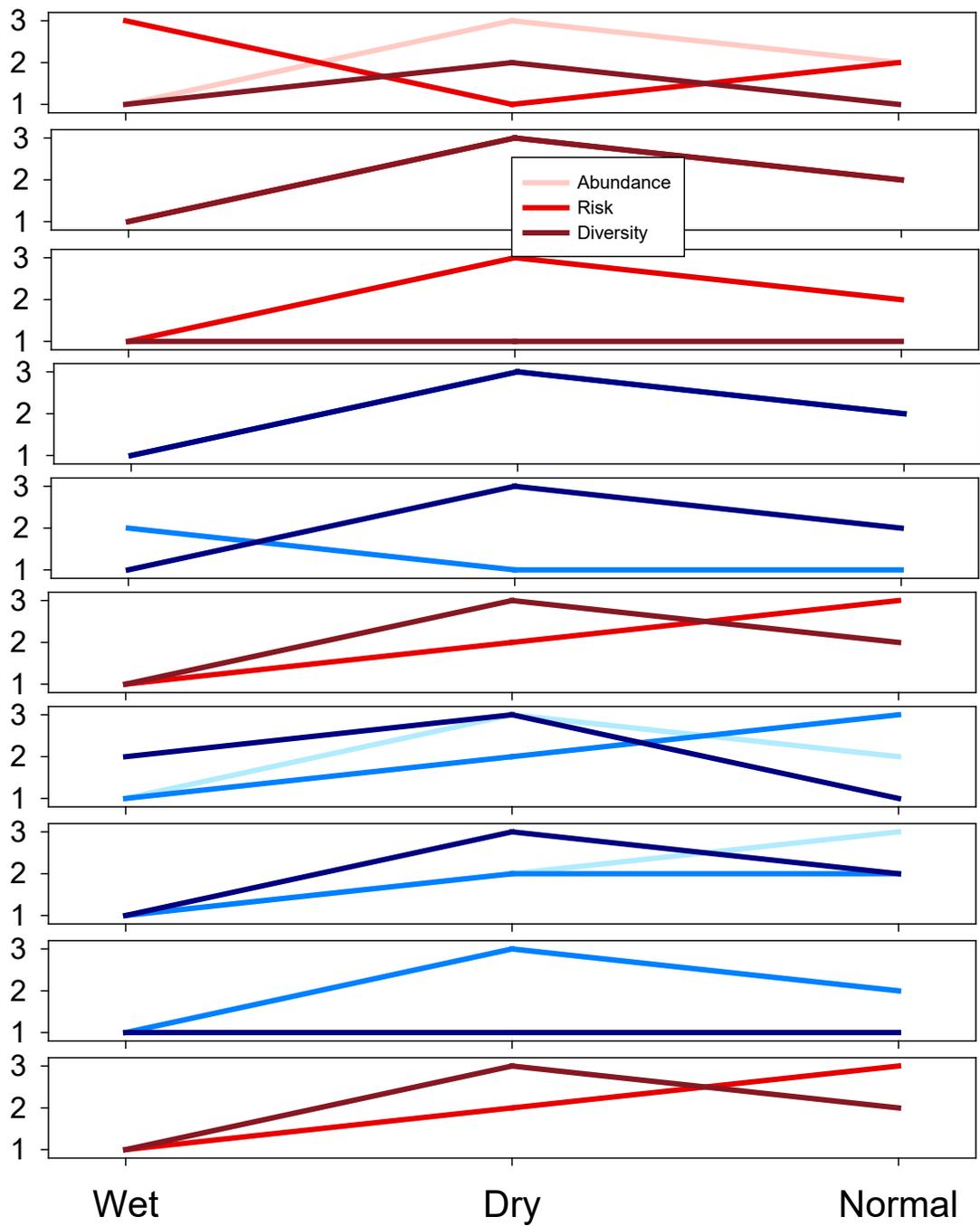


Figure 6. Graphical display of the rankings of wet, dry, and normal years from the EDT model for the same ten selected population-alternative combinations shown in Figure 4. The adverse effects rating is the color of the lines with Major as red and Minor as blue. The shading of the red and blue lines in each panel show the metric (light –abundance, medium – risk, and dark – diversity).

D. Going Forward

The multiple model approach being used by the Corps can provide valuable information on uncertainties that may lead to better decisions with higher confidence. The initial phases that ended with the public release of the Draft EIS represented a great effort, and the Corps has indicated that they plan to use these models in the future. In addition to any further use for the EIS, the multiple models can be very useful for the Biological Assessment, implementation decisions related to the project, monitoring design, and adaptive management. Beyond the recommendations made above, the ISAB offers some additional suggestions to consider going forward:

Suggestion 10: While the Corps has documentation of the models and their application for the Willamette Draft EIS, it is not clear how the models will be maintained (curated) and used for future analyses. Options are to: (1) repeat the same system of contracts and access to the modeling groups (often expensive and cumbersome), (2) simplify future applications by careful comparison of the models and predictions and eliminating one or more models because the information it provides is well represented by the other models, or (3) set up a technical advisory group (internal or external) that directs future application of the models.

Suggestion 11: Designate a small team to synthesize and evaluate the results for the EIS (i.e., a post-audit) to “learn” from this initial application of the multiple models. The team would examine the multiple modeling process to identify changes to improve future analyses. The team could explore issues such as loosening or tightening operating constraints, benefits of more years than the three from the supporting models used in the Draft EIS, aspects related to interpretation (why predictions occurred, adjusting the calculation of prediction variables), and how results are presented to various audiences.

Suggestion 12: Design how the multiple models would be used for future analyses in a strategic and tactical planning mode. Simulate possible analyses using old or mock (hypothetical) results to “test” the analysis plans and to maximize the usefulness of the analyses when applied to simulations for informing management later. Consider adding the uncertainty of predictions from individual models into the interpretation of predictions across the multiple models. Such estimates of prediction uncertainties are available from the individual model analyses though comparisons across models can be challenging because they were estimated differently. Another approach to ensure comparability is to prescribe the details of a sensitivity or uncertainty analyses (e.g., which inputs to vary and by how much) and implement that in all of the individual models for a common set of alternatives.

Suggestion 13: Convene a workshop with the modeling teams and various end users to leverage the knowledge accumulated by the modelers based on their completed single model analyses. The purpose of the workshop would be to inform the Corps more thoroughly about the models and their Willamette application than what is available in the existing modeling technical reports. For example, the modelers have accumulated a wealth of information on model behavior, important inputs, processes, and parameters, and likely have information on adjustments to the model that did not improve model performance and thus were not documented. It is important that the workshop not be about which model is better, but rather how to leverage the multiple models further. These types of workshops are often better led by a third party who is knowledgeable about modeling but not involved in the models being explored. Trust plays a major role in determining the effectiveness of these types of workshops.

Suggestion 14: Develop graphical and other methods for communicating the multiple model results. Effective techniques for illustrating complex model results are available (our figures are simple illustrations; see also Lewis et al. 2021), and the Corps may find science communication concepts for technical and general audiences useful (Schmolke et al. 2010; Chagaris et al. 2019; Peterman 2004; Welp et al. 2006; Nisbet and Scheufele 2009)

Suggestion 15: Plan and design sustainable protocols for code management, model inputs and outputs documentation and storage, and data exchange protocols. Key personnel will likely leave, and models will continue to be updated. Planning so that past analyses can be repeated and future analyses build on existing models ensures continuity. While the Willamette multiple model approach is much simpler than the massive climate change models and other multiple model examples, concepts from these well-studied and well-tested multiple model analyses could be helpful.

III. ISAB Review Comments on the Individual Models

A. Fish Benefits Workbook

Since the ISAB's review in 2014 ([ISAB 2014-3](#)) and the FBW's initial use, the FBW has undergone slight revisions with additional data added for some parts of the model. During our review, the Corps provided additional supporting documentation on inputs such as parameter values for Chinook and steelhead. The following updates have been added to the FBW since the 2014 ISAB review:

- Data on fish behavior, passage survival, route survival, migration, route efficiency, and structural collector performance with at least 10 additional reports.
- Project specific survival. Several reports with updated information.
- Dam passage efficiency. Several reports with updated information.

The ISAB recommends that the Corps incorporate updated information and changes in the procedures to the FBW user manual, which is dated 2014. The introduction states it is a living document, but changes since the original document are not apparent. For example, editorial changes suggested in the 2014 ISAB review have not been incorporated. A document that summarizes the new information in the additional reports over the years would be helpful. For example, a detailed bibliography with a summary for each report on FBW updates would be very useful. Some changes appear to be included in the parameter support information documents, (e.g., *Steelhead_FBW param support infor-102921.docx*), but few details are provided. Does the Corps store all documentation in a central website? Open access to that documentation would be useful.

One of the major concerns expressed in ISAB (2014) was the inability to assess the impacts of climate change because of the historical record used in the FBW only reflects past patterns of weather. The Corps included a qualitative assessment of potential effects of climate change in [DPEIS Appendix F: Qualitative Assessment of Climate Change](#). The assessment found little evidence in long-term trends (50 to 100 years, depending on location) for changes in observed peak streamflow hydrology in the Willamette Basin. It would be useful to compare the "envelope" of environmental and hydrological conditions of the historical record used to develop the FBW with the range of conditions projected for future climate change. However, several reports and the 2015 USACE Literature Review concluded that there is evidence to indicate that "temperature extremes in the Pacific Northwest show an increasing trend over the next century," and such increases are more pronounced for summer months during seasonal low flow. The literature review also found strong consensus that the frequency and intensity of extreme storm events will increase. The Corps concluded that action alternatives would

ameliorate some of the possible effects of climate change to a greater degree than the No Action Alternative. The Draft EIS indicates that additional measures could be taken if future observations or refined analyses indicate that the Preferred Alternative does not adequately ameliorate climate change effects. The DPEIS Appendix F provides qualitative and quantitative information that could be used to update the models to assess the consequences of future changes in environmental conditions and flow related to climate change.

B. Review of the University of British Columbia Models

1. Summary Comments

The UBC report evaluates the impact of the WVS EIS management alternatives on population trajectories of spring Chinook salmon and steelhead using separate life cycle models. The LCMs for both species are data-limited for several components. The UBC report mentions these data deficiencies in different places, but it would be more helpful to have all data gaps summarized in one place, followed by a discussion of these data gaps to help future modeling efforts.

The spring Chinook salmon LCM resembles the OSU and NOAA models in that parameter values are estimated for each component of the model separately as inputs, with some calibration for a few parameters. Finally, uncertainty is incorporated by drawing parameter values from a distribution for each parameter.

The pathway for model fitting would benefit from a clearer description. The model fitting combines likelihood-based, semi-Bayesian, and least-squares fits, rather than a uniform fitting procedure (a Bayesian approach would be preferred for most components). In general, the text is dense, the terminology is hard to follow, and possible errors in the descriptions of modeling procedures prevents a rigorous evaluation of how the LCM has been implemented. Many of the details for parameter estimation are found in Appendices, each of which are reviewed below.

A key issue is that not all sources of variation have been included, thus variation across simulation runs is likely underestimated. In addition, the 30-year planning horizon used to evaluate the alternatives can greatly influence some predictions, especially estimation of extinction risk (e.g., $P(QET)$). Also, there was little evaluation of the role played by the different supporting models in affecting model predictions. For example, there is indication (although not definitive without further evaluation) that the LCM results may be strongly dependent on the differences in survival estimated by the Corps' Fish Benefits Workbook (FBW; see brief FBW review above). One approach would be to compare the ranking of the alternatives by the UBC models to the ranking of the alternatives based solely on the FBW. There is also limited discussion on which performance measure is most appropriate if the rankings of the alternatives differ among the measures. The modelers did not attempt to create a summary measure.

A sensitivity analysis is presented, but it is difficult to compare across parameters because the results have not been standardized. Because no compensatory survival effects are modeled, the results showing that survival parameters are most influential are not surprising.

The LCM for steelhead is much simpler than the spring Chinook salmon LCM because only limited data are available. Similar to the spring Chinook salmon model, parameter values are estimated for each component of the model separately as inputs and a relatively simple calibration is done. Many details of the model are available in the report, but no summary diagram of the final LCM is presented to help the reader understand the model structure. Also, no information on model fitting is provided. For example, was a Bayesian model actually fit? Was an ad-hoc minimization procedure used? How was uncertainty in the parameter estimates determined?

Because of a lack of data, most of the parameters in the steelhead LCM are assumed (refer to Appendix B). Unlike the spring Chinook model, the UBC steelhead model incorporates no variability in these parameters over simulation runs. This implies that the variability in output across simulation runs will be underestimated. A table that explicitly outlines the sources of variation should be provided.

The analysis of sensitivity to marine survival and dam passage is useful and needs more explanation and interpretation. The listing of tables is a start and should be augmented with narrative guidance for the reader. Many tables are included that provide useful information but are difficult for readers to digest; we recommend that graphical summaries would be beneficial and the authors could provide the numerous detailed tables in appendices.

2. ISAB Answers to the Review Questions

The ISAB has developed the following sets of questions for the review based on questions provided by the Corps and its internal discussions.

1. Model structure and development

- a. *Is the model structure appropriate for evaluating fish responses for the EIS alternatives?*
 - i. *Are the relevant processes (mortality, growth, reproduction, movement) included and are they represented in a way that encompasses the changes within and among the alternatives?*

Both LCMs include the major processes involved in a population trajectory, and the models seem reasonable. How much the results depend on the outputs for each alternative derived from the FBW, which is treated as a “black box,” needs further investigation. Consequently, any structural problems within the FBW or inaccurate inputs to the FBW will create inaccurate dam passage and survival estimates, which can then potentially drive predictions of the LCMs.

- ii. Are the key functional forms (e.g., shapes, dependence on hydrological and environmental variables) sufficiently realistic?*

In general, the functional forms (submodels) are well reasoned; the ISAB provides both general and specific comments for each component in our detailed review.

- iii. What aspects of the model are stochastic?*

While it is impossible for a model to include all sources of variation, the spring Chinook salmon and the steelhead LCMs may be missing some key contributors. For example, neither model accounts for demographic stochasticity and many parameters are represented as fixed. Consequently, the variability in the response measures may be underestimated, which would greatly affect the estimated P(QET).

- b. Are the models formulated to assess fish responses at the appropriate temporal and spatial scales that both capture the effects of the alternatives and generate outcomes relevant to management scales?*

The temporal resolution is a “year.” The FBW output for dam passage and survival are available at finer temporal resolutions and are combined for each year. This is a sensible scale for the LCM developed here. The management horizon was 30-years plus a 2- or 5-year “burn-in” period.

- c. Are the estimates used for model input parameters (hydrological, environmental, processes) reasonable and scientifically defensible? What data are used to estimate or confirm model parameter values?*

Model parameters were derived from studies in the basin, data on nearby populations, published studies, or expert opinion. Many estimates of parameters are based on Zabel et al. (2015), similar to what was done in the NOAA models. Finally, some (limited) calibration to a (small) set of current data is used to inform some parameters. Calibration to additional years with different hydrological conditions would increase the confidence in the calibration.

- d. What sources of variation that affect population dynamics and responses have been accounted for?*

Environmental variability is modeled by using a set of historical water years and bootstrapping values from this set. Uncertainty in some parameter values is accounted for by sampling from a distribution centered about the parameter estimate with suitable variability. However, some parameter values are fixed (e.g., ocean survival). The FBW output for each alternative does not

incorporate stochasticity. Consequently, overall variation in population dynamics is likely underestimated.

- e. Have the researchers adequately evaluated the sensitivity of the predictions to uncertainties in various sources of data?*

There is some discussion of sensitivity to data in the report, but the results are difficult to interpret because the results were not standardized to a common scale. Because compensatory processes after egg laying were not modeled, it is expected that survival parameters that act in a density-independent manner greatly influence the results.

- f. How sensitive are predictions to uncertainties in hydrological conditions, environmental variation, and model parameters?*

The final results appear to depend strongly on how the FBW predicts changes in dam passage and survival. The FBW does not include any stochasticity.

Potential effects of climate change on the alternatives were not analyzed with the spring Chinook and steelhead models. Consequently, comparisons based on historical water-years may not reflect future outcomes of the alternatives, and results should be interpreted carefully.

- g. Are subcomponent models adequately integrated into the larger model?*

For both species, a single model integrated the subcomponents in a standard fashion.

2. Key assumptions and limitations

- a. Are the model assumptions and limitations sufficiently documented? What are the key assumptions, strengths, and limitations?*

The assumptions for each part of the LCM are summarized and briefly discussed. The FBW is treated as a “black box,” and no assessment of the reality of its output or internal assumptions is made. Because the differences in the FBW output under each alternative are potentially major drivers of the VSP scores, the internal assumptions and structure of the FBW are crucial. The FBW is very briefly reviewed in this document and has been reviewed in detail by the ISAB elsewhere ([ISAB 2014-3](#)).

- b. Is the reasoning, rationale, and evidence for the key simplifying assumptions (e.g., no interactions with other species) reasonable and adequately documented?*

This is not discussed in the document.

3. Model validation and verification

- a. *Does the model documentation adequately describe testing steps utilized during model development (i.e., consistency check, sensitivity analyses, calibration, validation)?*

The document is mostly silent on these issues for each sub-component. Data to calibrate the LCMs were limited in their coverage of years, and a small data set is used for calibration to modify a few parameters to get what appear to be reasonable values for the NAA. The spring Chinook salmon model was calibrated against historical natural-origin returns and age structure for each subbasin population, varying several parameters. For steelhead, only two parameters were adjusted to calibrate the model against adult returns for a single location (Foster Dam). These differences reflect the limited available data for steelhead.

- b. *Have the researchers adequately assessed the fit of model to current data? Have the researchers assessed the fit of the model to new data (e.g., use it to predict years not used in fitting and compare)?*

Assessment of model performance by comparison to data is limited by the availability of the data. Thus, there is more confidence in using model predictions to compare relative responses among alternatives than using model predictions of abundance on an absolute scale.

- c. *Has the model programming or system components been tested for computational correctness and associated errors? If not, what is the potential for errors to occur?*

This is not discussed in this report. However, the ISAB identified possible areas of concern about model implementation. These are described in our detailed comments.

- d. *Has sensitivity and uncertainty analyses been applied to the model and the results used effectively to identify sensitive parameters and quantify variability around model predictions?*

This report contains some information on a sensitivity analysis, but it is difficult to digest and needs more discussion and interpretation.

4. Data gaps

- a. *For which part of the model is data strongest and weakest? Is the model sensitive to these data gaps?*

Models for both species lack data for a number of components. The report mentions these data deficiencies, but that information is often described in different sections of the text (e.g., page

22 on reservoir survival, page 24 on TDG mortality) rather than summarized in one place. A discussion of key data gaps is needed to inform future modeling efforts. Determining the sensitivity of the models to specific data gaps can be systematically addressed with sensitivity and uncertainty analyses.

Generally, the lack of steelhead data specific to the Willamette basin leads to relatively high uncertainty in the WVS-specific predictions for steelhead. Obtaining site-specific information for the critical inputs used to develop parameters would reduce model uncertainty.

b. If so, then what type of research program is needed?

Development of life history parameters that are based more from the Willamette River basin would provide better inputs for this and other models of Willamette salmonids.

5. Model output

a. Can the outputs of the model be used to rigorously compare and contrast different management actions?

The authors present detailed results for each performance measure, but they do not identify which performance measure is most appropriate, do not discuss whether the different performance measures rank the alternatives in the same way, and do not provide an overall summary measure. As noted in our detailed comments, results may be highly dependent on FBW outputs.

i. How reliable are the comparisons of the effects of the alternatives (e.g., can be used only to rank alternatives; can be used in cost-benefit analyses)?

The key drivers of the results may be highly influenced by the differences in dam passage efficiency and survival from the FBW. Calibration was used to match model outputs with the data was limited by data availability. Not all forms of stochasticity were captured in the LCMs. Consequently, the results are strongest (highest confidence) when viewed as relative changes; interpretation of model results as abundances should be done carefully and with caution. Model predictions may be most robust when used within a ranking scheme.

ii. Is the output stochastic so questions about quasi-extinction can be answered?

The output is stochastic, but with stochasticity lacking in some key inputs (e.g., FBW passage effects) and the lack of strong calibration to past data imply that inferences about quasi-extinction are likely highly uncertain. As well, the 30-year planning horizon (plus a 2- or 5-year “burn-in” period) used for the models may be insufficient to reliably estimate P(QET).

iii. *Do the researchers distinguish whether differences in predicted responses among alternatives are biologically meaningful?*

No.

b. *What specific metrics or types of output are provided to the user? Are these outputs sufficient for evaluating the EIS alternatives? Would additional types of output be useful?*

Many figures and tables are presented, and further discussion is needed on how to interpret the results. No summary measure of all the performance measures is computed. Some figures are presented on the underlying metrics, but further discussion would be beneficial.

c. *How far outside of the envelope of data used to develop the model will climate change take us? Do we have good information on how the model will perform outside of its development based on historical conditions?*

No information is presented on impacts of climate change or regarding model performance outside the range of data for which the models were calibrated.

6. Model improvement

a. *What improvements can be made to address any critical shortcomings of the models?*

The ISAB identified a number of areas where the model fitting may require reconsideration to correct potential errors (see detailed comments below).

b. *Where should future model development be focused?*

The biggest improvements would come not from modifying the models *per se*, but rather by improving the model parameters and possibly the FBW inputs to better represent natural variability and improving and expanding the data available for calibration.

3. ISAB comments on UBC Report Chapters

Review of Chapter 1. Introduction

The title of the report is broad, encompassing “modelling to evaluate alternative... measures for wild spring Chinook salmon and winter steelhead” but on p. 4 it states that “Our modelling focused on assessing the outcomes of implementing downstream passage alternatives...” Perhaps this important distinction and more focused scope should be made clear in the title.

Overall, section 1.1 gives a good, readable overview of the model structure, without getting lost in implementation details. However, the introductory chapter then becomes very detailed, with complex tables, many acronyms, jargon about burn-in periods, and such. A flow chart would greatly help readers understand how the models work, especially given the focus on juvenile passage for evaluating alternatives. The modelers should emphasize the movements and survival rates of fish at different stages, relative to alternatives. There must be a way to present this more holistically, before getting to the myriad details. The UBC report would benefit from an executive summary to allow readers to understand the results, key findings, and limitations without needing to read the entire document.

Unlike the NOAA models, only a 30-year management horizon is used, which was dictated by the Corps. However, this is likely too short to compute a reliable long-term value of P(QET). The mean abundance is computed over years 16-30 (15 years) which again is short when compared to the length of external cycles (5-10 years long). Similarly, productivity is computed over the first 5 years, which should be better justified given the relatively long life-cycles of the species.

The summary of the limitations and differences between the UBC models and NOAA models is useful and should be considered in the Corps' multi-model comparison. The summary of the FBW provides useful information, but it is extensive and is probably most helpful for someone not familiar with the FBW.

Specific comments

p. 2. The authors bootstrapped records of historical year flow and river water temperatures. This might obscure any autocorrelation across years induced by external drivers such as ENSO or PDO?

p. 2. The authors use a quasi-Bayesian approach with a minimizer but do not use a fully Bayesian (e.g., MCMC) approach to fully define posterior distributions.

p. 5. "... it was assumed that the returning adults would self-sort ..." into HOR and NOR. This questionable assumption is not discussed later and has implications for hatchery-wild fitness effects.

p. 5. The background context for outplanting was helpful for getting context on the baseline and assumptions.

p. 7. Productivity measures used a geometric mean for one performance measure (R/S) but arithmetic means for SAR and smolt survival. Either a geometric mean should be used for all three measures or the difference should be justified.

p. 8. Productivity metrics are calculated over years 1-5. This is justified because they would be less informative if calculated near equilibrium. This then assumes that current abundance is NOT near equilibrium – is there evidence for this?

p. 8. A given run of smolts returns in a number of future years. Yet, the description on page 8 seems to only use adults versus smolts five years previous? It is unclear if this is a combination of multiple smolt runs. This should be clarified or corrected.

p. 8. Not only can 15-years be short to evaluate of P(QET), but it is chosen as being “when the population was at or near equilibrium.” This needs to be clarified. Extinction risk is often influenced by dynamics during population transitions, and only focusing on stable periods can bias (underestimate or overestimate) risk. It would be helpful to compute P(QET) over various time spans (e.g., first 10 years, last 15 years, all 30 years) entire to assess the robustness of the estimate. The 30-year time frame is also inconsistent with Technical Review Team (TRT) definitions, which look at a 100-yr time frame.

p. 8. The report states that “Recruits were measured as returning NOR spawners...” and this seems to neglect fisheries effect, which can vary over time, as well as the possibility of natural variability in marine survival. At a minimum, this should be made explicit, and any implications of the assumption noted as well. This model, as well as the other models, ignores sources for CWT-based estimates of brood year survival and ocean exploitation for Willamette Chinook. These should be referenced, if not used (e.g., TCCHINOOK(23)-01-Supplementary Material, TCCHINOOK(23-01); <https://www.psc.org/publications/technical-reports/technical-committee-reports/chinook/>). Recent ocean exploitations have averaged 8%, though other lower Columbia stocks have shown decadal variability.

p. 9. Different burn-in periods were used for Chinook and steelhead. The justifications seem reasonable, but it is hard to keep track of all of these assumptions and differences between the two LCM models. Tables 1.3.1 and 1.3.2 attempt to summarize the performance metrics (and methods to some extent) as efficiently as possible, and are probably adequate, and it would be helpful to summarize the assumptions in a table or box (similar to the table in section 2.6).

p. 9. The discussion of burn-in periods is confusing. Earlier, the text indicated that model runs were 30-years. Does this mean that runs were 35 (32) years, with model year 1 being actually

simulation year 6 (Chinook) or 3 (steelhead)? The text would benefit from better explanation of how years line-up and better justification for the burn-in lengths.

Editorial

p. 2. "... drawn from prior distributions with central tendency and prior modes..." What is meant by "with a central tendency"? What is a "prior mode"? Here, the authors' description has an intended meaning that is not obvious to the ISAB reviewers. The ISAB recommends providing additional description in the introduction or more likely in the appropriate analytical section.

Review of Chapter 2. Spring Chinook life cycle model

Summary

The spring Chinook life cycle model is similar to the OSU and NOAA models in that model parameters are estimated for each component of the model separately as model inputs. Some calibration is done for a limited number of parameters. Simulated trajectories are then created with uncertainty in the parameters being incorporated by "sampling" parameter values from assumed distributions for each parameter.

Many of the parameters here are taken from Zabel et al. (2015), so any concerns expressed about the NOAA model would also apply here. The Zabel report is over 400 pages, and one has to carefully review it to find needed information. Figure 2.3.1 is a good illustration of the pathways, but the report would benefit from estimates of the average or median values for the major decision points (e.g., proportions surviving to each age and maturing or not, etc.). A simplified version of Table A.6 would be very helpful.

The pathway for model fitting would benefit from a more transparent description. As presented, model fitting is a combination of likelihood-based, semi-Bayesian, and least-squares fits, rather than a uniform fitting procedure (a Bayesian approach would be preferred for most components). For example, in Section 2.5, a least-squared "likelihood" function is presented, along with what appear to be prior distributions, but then a combined objective function is minimized which is neither a likelihood nor a Bayesian analysis. In general, the text is very dense with terminology, which is hard to follow, and possible errors in the descriptions of modeling procedures raise questions of how the LCM was implemented.

The appendices contain many of the details about how parameter estimates were obtained. The modelers should carefully review and update information in the appendices.

It is not clear if all important sources of variation have been included. For example, it appears that no demographic stochasticity has been included. As a result, variation across simulation runs is likely understated and P(QET) underestimated. An explicit table of the sources of variation included and excluded would be helpful. In particular, while environmental variation is included to some degree (3 different years are simulated), no autocorrelation is included (in contrast to the steelhead model). Temporal autocorrelation is a fundamental property of environmental variation (e.g., van der Sleen et al. 2022), and ignoring it potentially biases population analyses. In particular, ignoring positive autocorrelation will result in low estimates of extinction risk, so the P(QET) results reported here may be significantly biased.

The authors provide a list of model assumptions (Table 2.6). The third column indicates how the assumption was assessed. Some of these assessments need further discussion. For example, the table indicates that the modelers assessed whether egg-fry production follows a BH model, which implies that other models for this stage were considered. Using a distribution to model uncertainty in the parameters of the BH or employing a sensitivity analysis does not assess this assumption sufficiently.

In some cases, these assumptions should be re-visited and the model adjusted if appropriate. For example, carcass surveys do not accurately represent true age composition. Zhou (2002) reported a very substantial age (and some sex) bias in Chinook salmon. The assumption of no straying is not justified; its importance can be assessed. Spring running adults and stream type juveniles tend to have lower straying rates than fall/ocean type Chinook (Westley et al. 2013, and this might be mentioned as justification for the assumption.

One key assumption made in construction of the LCMs is the lack of compensatory effects in other stages. This has consequences on the sensitivity analysis (see above), and likely explains why the results seem to be driven by the changes in DPE (dam passage efficiency) and DPS (dam passage survival) as modeled by the FBW.

It is unclear if the estimated annual deviations in marine survival for the period of data availability are used in the projections. Are they treated as random effects?

The results from the model fitting are summarized for each basin using a Table (e.g., Table 2.7.2) and a graph of the range of simulation outputs (e.g., Figure 2.7.3). Most of the plots report the range of outcomes as “confidence intervals.” These are not confidence intervals but are prediction intervals. An error-bar is used to represent the range between the 2.5th and 97.5th percentiles, but this format does not distinguish between outcomes where the simulation results follow a distribution that is close to uniform or is peaked around the

mean/median. Violin plots of the results from the simulation would be a better format to display the results. Note that the format for displaying the results changes for the steelhead results – a common format should be used for both species.

The LCMs do not include some potentially important sources of variation (e.g., demographic stochasticity is not included and many parameters are fixed at specific values), and so the ranges in response measures likely are too small. Are the differences among the alternatives so large that this extra variation is moot?

The LCM results may be heavily dependent on the differences in survival measured by the FBW. Does the ranking presented here match the ranking as determined by the FBW? If not, why not?

Do all the performance measures rank the alternatives in the same way? If not, why? The authors need to create a summary measure that combines the response measures in a sensible fashion as was done in the NOAA results. This would help the reader interpret the results.

There are figures, tables, and appendices with much detail, but it would be very helpful to have a simple presentation of some of the key pieces of information in the main body of the report. For example, estimated fecundity of wild and hatchery adults, average egg to fry survival, natural and fishing mortality at sea, etc. This compact presentation of essential values (admittedly, neglecting density-dependence, variation, and so forth) would greatly assist readers, who otherwise will have to search for the information. At the very least, direct the reader to the appropriate section in the appendices.

Finally, a sensitivity analysis was performed that reports the mean natural origin return (NOR) abundance after prespawning mortality in simulation years 26-30 by modifying one parameter at a time and keeping all other parameters fixed. Hence, the LCM is now treated as a deterministic model for the sensitivity analysis, which tends to exaggerate the impact of changing a single parameter. The ranges are not standardized for different parameters, so the reader cannot easily rank parameters based on sensitivity. Converting changes in NOR to marginal change per change in the parameter would aid in interpretation. We recommend that the slope of the curves in the figures in this section be evaluated at the current mean value of the parameter, to convert all sensitivities into a common scale. The authors conclude that the early ocean survival has the most influence on the mean NOR in years 26-30 based on a visual inspection of the graphs.

This sensitivity analysis implicitly assumes no compensatory responses elsewhere in the LCM – indeed none were included in the LCM, so the results obtained are expected.

Specific comments

p. 16. Downstream survival and smolt-adult survival were estimated using a BCJS method, which appears to have implementation concerns – see our comments about Appendices C, D, and J.

p. 18. Density-dependence is only implemented as a Beverton-Holt function at one life stage (egg-fry), which could be appropriate because (as they point out) BH functions multiplied across different life stages result in an overall BH function for the whole life cycle. However, this would only work if the parameters were estimated during population calibration (fitting). Here, they estimate the slope parameter (a) as part of fitting, but the capacity parameter (b) is estimated from spawner habitat capacity only. Thus, if there are habitat bottlenecks at other life stages, the overall capacity will be overestimated (density-dependence underestimated). They also state an assumption that current spawners are well below capacity, so there will be little density dependence without significant downstream passage improvements. This assumption needs additional explanation and justification.

p. 19. The distinction between Chinook fry and parr migrants (e.g., Figure 2.2.1 and 2.2.2) refers to Monzyk et al. and Romer et al. papers as listed in their reference section, but these seem questionable. In 2.2.2, are fry really emerging in August? Similarly, in 2.2.1, some years show clear separation at around 60 mm but other years do not. This apparent cutoff value is 15 mm higher than the 45 mm used by Zimmerman et al. (2015), and Anderson and Topping (2018). Chinook fry are not emerging at 60 mm, so this cutoff clearly includes some growth. Perhaps the overall effect of this is small, but some acknowledgement should be made of these differences.

With respect to ocean survival and age at return, it would be very helpful to have a simple table and information in the text indicating the proportions of adults expected to return at different ages for each of the fry, parr, and yearling smolt categories. There are many details and factors, and having these general values will be helpful. For example, could Figure 2.3.1 be expanded in size to have the proportions shown in each box (e.g., mortality rates, maturation schedule, etc.)? This could be shown also for the other life history pathways (fry and parr migrants).

In 2.1.2, egg-fry survival rates and density dependence are mentioned. What estimates of fecundity were used, and what were the survival rates at low density? This kind of basic information would be helpful to include in the text.

p. 24. In 2.2.4, the indication that pre-spawning mortality is typically “all or nothing” is consistent with some studies outside the basin (e.g., Barnett et al. 2020) as well. Importantly, the net effect of this assumption is probably not large, depending on the distribution of egg retention values. Might some comment to this effect be made?

p. 24. Section 2.2.5 – Total dissolved gas (TDG) mortality is not included in the model because data do not exist to support it. TDG exposure will definitely be different among the scenarios. The results would have been stronger if they had included a call for more data about TDG-related mortality and at least a statement of the assumption on whether including it would affect their findings regarding the alternatives.³

p. 24. The report stated, “We assumed that posterior estimates reflected long-term average values... uncertainty accounted for in the posterior distribution.” There are two sources of variability – uncertainty in the estimates and year-to-year variation around the long-term average values. Exactly what do the posterior estimates/distributions cover? Both cannot be covered by a single distribution.

p. 27. They discuss shortcomings of the marine survival estimates from the Pacific Salmon Commission Chinook Technical Committee (CTC), and further say that CWT data were not available for model fitting. All CWT data are available in the Regional Mark Information System (RMIS Standard Reporting, rmis.org) as well as published reports (e.g., <https://www.psc.org/publications/technical-reports/technical-committee-reports/chinook/>). These data have been analyzed in several publications (e.g. CTC reports, Welch et al. 2021). It is not clear what specific data are not available.

p. 28-29. The discussion of PSM, temperature, and the proportion of hatchery-origin spawners (pHOS), and their justification for downweighting pHOS, is inconsistent. Because the initial model has a poor fit to new data, they downweight one of the previously determined significant variables based on an assumed (not observed) above/below dam effect on pHOS-survival relationship and the claim that no mechanistic explanation of the pHOS effect exists; elsewhere, they have accepted non-mechanistic relationships.

p. 31. It is unclear why the model uses a triangular relationship for HOR spawning success. The modelers need to explain the basis for the relationship.

³ Note that the Corps used other models to assess water quality effects of the management alternatives.

p. 31. The model does not separately track HOR and NOR fish through the lifecycle. Rather, it only separates them as adult spawners and eggs, based on outplanting rules. Are HOR fish irrelevant in the rest of the life cycle (e.g., no food competition or disease interactions)? This seems unlikely.

p. 31. In 2.4.3 reference is made (p. 31) to the work of Chilcote et al. (2011, 2013) indicating that “a population composed entirely of hatchery-origin fish predicted [to have intrinsic performance] to be only 6% that of one composed entirely of wild fish.” This might be seen as extreme, as other studies have reported greater similarity in realized reproductive success of wild and hatchery fish. This is a very complicated topic, with many studies on Chinook salmon, steelhead, and (to a lesser extent) other species. We do not need a review here, but some consideration of the complexities would be good, especially if such an extreme value is used. Whether the hatchery fish are of local or non-local origin matters greatly, for example. The text that follows moderates that citation (p. 31) well. It might be helpful to indicate how sensitive the results are to the mean and range of values used. The assumption that HOR will only spawn with each other seems odd. Why was this assumed and how consequential is this assumption? Given salmonid mating systems, it seems unlikely, unless spatial or temporal isolation is extreme.

p. 32-33.

Equations 2.5-1 and 2.5-2 claim to be the likelihood components fitted using least squares. A least-squares measure is a log-likelihood kernel assuming a normal distribution for the data around the mean. Is this a sensible distribution based on observation and biology? For example, it appears that a log-normal distribution may be a more appropriate distribution for NOR. Assuming a normal distribution for the p parameters - which are proportions - is not realistic and may bias the estimates. In both cases, an SD appears in the denominator and is assumed to be usually 1. What does this mean, especially when in the text under the equations, the SD is “assumed to 1 or the NOR time series” and “assumed SD for the proportion-at-age time series (usually 1.0)”, neither of which are explicitly justified as realistic.

These equations are not independent of units (e.g., changing p from fractions to percent should give the same fit, but it will not).

Equation 2.5-2 using the proportions at age are summed across the ages, but these are not independent parameters (the proportions must sum to 1). This restriction is ignored in Equation 2.5-2.

Equations 2.5-3 to 2.5-6 are stated to be prior distributions, but none of these equations are distributions. What are these equations supposed to represent? It appears that perhaps some sort of normal distribution appears to be involved, but these equations do not specify this. Some objective functions (Equation 2.5-7) involving likelihood functions and priors were minimized. However, a Bayesian approach does not minimize an objective function, so it is not obvious what type of fitting was done. Did the authors conduct a likelihood analysis, in which case the log-likelihood can only be functions of observed data and parameters and not latent variables such as random effects, or a Bayesian analysis, in which case the likelihood component can also include latent variables and the MCMC integrates over them? In the latter, an “objective” function is not maximized, so the “priors” now look like penalty terms but there is no weighting to the penalty terms. It appears the authors did something ad hoc, which should be justified statistically.

p. 33. It appears that deviations in marine survival are identical for yearling and subyearling migrants. Is this a justifiable assumption, and is there any documentation supporting it? Again, there are data to evaluate this. Assuming that subyearling migrants make it to salt water in their first season, their survival would be expected to be much less than for yearling migrants (e.g., CTC reports above and Welch et al. 2021).

p. 33. “Bayesian priors were included ... when the initial model fit produced a zero value...” Were no priors included for other parameters, or were they default non-informative priors, or something else? Is this not a Bayesian analysis?

p. 34. Table 2.5.1 indicates that it shows the “prior mean and prior SD” values used in the calibration process. Are these fixed and not estimated? What exactly are the SD parameters and the mu parameter? What is a “prior mean” and “prior SD”? Do the authors imply these are the mean and SD of the prior distribution for these parameters?

p. 40. According to the UBC document, “Under the NAA, the model reached an equilibrium within five years.” What type of equilibrium is meant here? This only makes sense when deterministic predictions are used and demographic stochasticity and variation in parameter values are ignored. This statement appears to refer to Figure 2.7.2 where “equilibrium” appears to refer to the median over the 10,000 simulations.

p. 41. Figure 2.7.1. The model predicts NOR and is presented as a line joining point estimates across years. To present observations for seven years as a single point value hides annual variation. Are the observed points means, totals, other?

p. 42. There is not nearly enough variation over time in the NOR. Figure 2.7.2 purports to show 95% confidence intervals for the NOR, but this seems incorrect unless something else is intended. With 10,000 simulations, the median will be known with almost certainty. It appears that the dashed lines are 2.5th and 97.5th percentiles over the simulations.

p. 44. As we indicate above, these are not confidence intervals because with 10,000 simulations, the uncertainty about the median will essentially be 0. These appear to be 2.5th and 97.5th percentiles over the simulations.

p. 45. There appears to be discrepancies between this figure and Figure 2.7.2. Under the NAA, the NOR spawners range from about 1000 to under 2500. However, Figure 2.7.2. shows a range of NOR spawners to approximately 4000.

p. 47. Our previous comments about “confidence intervals” apply here.

p. 49. The ISAB was confused by the statement that “there was almost no uncertainty in DPE.” First, what does “almost no” mean? Second, how is there almost no uncertainty in DPE given the uncertainty there is in outputs from the FBW that serve as inputs to this model? On page 62, the report states that the DPE varies among the dams and alternative, but the authors do not really explain why.

p. 51. Our previous comments about “confidence intervals” apply here.

Editorial

p. 34. Table 2.5.1 What do the dashes in the table mean? This should be indicated.

p. 44. “Error bars are not shown for P<QET, owing to its binary outcomes.” This needs rewording – a binary outcome can still have a confidence interval computed. However, as noted previously, these are not confidence intervals.

p. 50. What is the Y-axis? A count or density value? The legend says that these are histograms but the Y-axis appears to be “density” since the area under the curve appears to be 1? Moreover, the scale of the Y-axis varies among plots to enhance the patterns and distributions. This should be noted and explained. This comment applies to analogous plots in subsequent sections as well.

p. 51. “Uncertainty in DPE and DPS.” However, the FBW is deterministic and has no uncertainty. What is being reported here? Variability?

Review of Chapter 3. Winter Steelhead LCM

This LCM is much simpler than the Chinook LCM because very few data are available.

Most of the parameters are assumed to be known (refer to Appendix B), and unlike the Chinook model, no variability in these parameters over simulation runs is incorporated. This implies that the variability in output across simulation runs will be understated. The report should include a table that explicitly outlines the sources of variation included or excluded.

There are only a few parameters to be estimated (Section 3.3.1). Equation 3.3-1 purports to be the logarithm of the product of prior distributions on parameters with many of the variances assumed to be known. Some parameters are not defined in Appendix B (e.g., σ_F). Equation 3.3-2 purports to be a log-likelihood for the process but actually is a least squares (assuming a log-normal distribution) and does not integrate over random (unobservable) effects and so is not a proper likelihood. It is also missing a term for σ_{SHL} .

The text says marine survival deviates have a log-normal prior; this is not consistent with eq. 3.3-1 (and eq. F-1) unless d_y is the deviation in $\log(\text{survival})$.

No information on model fitting is provided. For example, was a Bayesian model actually fit? Was an ad-hoc minimization procedure used? How was uncertainty in the parameter estimates determined? The first paragraph of 3.5.1 indicates that some parameters have standard errors computed, so it appears that some sort of likelihood fit is done; but then it is not clear why prior distributions are used.

Future projections (predictions?) do not appear to incorporate uncertainty in the parameter estimates, so it is not clear how variation in Section 3.5 is obtained? Appendix B appears to suggest that the DPE and DPS are bootstrapped from past values. Table 3.4 seems to indicate that ϕ_f is sampled from a posterior from the model fit. This needs to be clarified.

Many tables are presented, but these are difficult to digest without further explanation being provided. Some graphical summaries are desirable because it is hard for a reader to digest numerous tables of computer output. We suggest the authors provide the tables of underlying data for graphical summaries in a supplemental output section or appendix and refer to these sections in the figure captions. Figures similar to those for Chinook would be more useful than tables.

It is unclear why the presentation of results for steelhead is different from that for Chinook. Ordering of alternatives in graphs and tables differs from that of Chinook, and different abbreviations are used for alternatives in figures compared to tables and to Chinook results. Format and style should be consistent within the report. We recommend adopting a consistent presentation to help the reader. Violin plots may be better than box plots here.

The analysis of sensitivity to marine survival and dam passage is useful but needs more explanation and interpretation rather than just presenting a list of tables without guidance to the reader. The sensitivity analysis uses only three points for each parameter. Multiple values and showing a curve, similar to the analysis for Chinook, would be more informative

Specific Comments

p. 100. Refer to the text about the three stages for survival: The terminology here is confusing, with “freshwater” referring to only a portion of the true freshwater life history, “smolts” being applied to fish above the dams, and “marine” including upstream migration. Smolt survival is the survival of smolts from when they are tagged until they return to spawn. This includes mortality during both the downstream and upstream (adult) portions of the life history. Marine survival, which is rarely measured unless smolts are tagged near tidewater, is the survival of young salmon from the time they enter salt water to their return to salt water. It is important to have estimates of fishing mortality in freshwater (smolt survival only) and the ocean (for both smolt and marine) to be able to interpret survival trends, and such measurements usually require coded-wire tags.

p. 100. The use of a single BH function for egg-smolt survival is justified by evidence of density dependent freshwater survival coupled with density-independent marine survival for a single stock. The terms “freshwater” and “marine” as used here do not completely align with the usage in the sources cited. Potential density-dependent effects when multiple species or stocks are present are ignored. As for the Chinook model, this assumption could lead to overestimating habitat capacity across the entire life cycle.

p. 101. The assumption that all smolts are Age 2 is questionable, as smolt ages vary in response to environmental conditions. The data they cite (Section 3.2.1, p. 103) suggests that 81 – 84 % of smolts are Age 2. Is this close enough to 100% for the purposes of this model? This assumption needs to be justified.

p. 101. The authors state that “We incorporated variable marine survival by using an index of marine survival rate, which varies with an apparent approximate decadal periodicity...” Were marine survival and age composition for spring Chinook fixed over the years? If so, the authors

should explain why the species were treated differently. If not, the representation of marine survival for both species could be made clearer (e.g., Appendix Table D.5, on p. 225). Both the Chinook and steelhead models include variable marine survival, but steelhead includes autocorrelation while Chinook does not. The report should explain this more thoroughly and provide supporting references.

p. 112. In applying the model calibrated for Foster Dam to populations above Detroit and Green Peter dams, there is no mention of how density-dependence is handled. Were independent parr capacity estimates made for the two additional dams?⁴

p. 114. Section 3.4. As with the corresponding section for Chinook, the last column is not an assessment of the assumption.

Regarding the assumption of density dependence only at a single life stage for a single population, the capacity parameter will be too high if there are bottlenecks at other life stages. This is mentioned in discussion under potential extensions and should be included in the table.

p. 118. More statistics on the model fit would be useful, as well as plots of the priors and posteriors of fitted parameters.

p. 119. The list of performance metrics in Tables 3.5.1-4 is different from that for Chinook, and includes survival components (marine survival, DPS, DPE) that appear to be inputs, not performance measures.

Each table and figure should clearly indicate the species under consideration. The chapters are separate, but they are so long that one can get lost.

p. 126 ff. For Green Peter Dam, results for NAA and A4 are missing without explanation.

Editorial

P.100. The model is described here and in Appendix B. It would be helpful to create similar figures for the steelhead LCM as was done for the Chinook LCM (e.g., similar to Figure 2.1.1). The modelers presented such a figure in an earlier PowerPoint presentation, and it should be included here.

⁴ This might be answered in Appendix B.

p. 101. φ_F “is assumed to be density dependent” contradicts the next paragraph (and the equations) where φ_F is a single survival parameter, not a function. The authors probably meant to say “smolt production is assumed to be density dependent.”

p. 116. What is the “base case parameterization” mentioned in the section heading – this has not yet been defined.

p. 118. Figure 3.5.1. “The confidence intervals of the fitted counts...” How were these obtained?

p. 143. Figure 3.6.6 is a repeat of the Green Peter figure 3.6.4, rather than results for Detroit.

Review of Chapter 4. Discussion

The authors should summarize the many performance measures presented earlier or discuss which is most important and least important. For example, the authors state

“ (p. 153) ...the probability of exceeding QETs for different EIS alternatives were markedly different when in contrast the mean values for the FBW outputs were practically the same between the EIS alternatives.”

Which performance measure is a better indicator of stock status under what conditions (e.g., how is P(QET) better/worse than pHOS) etc.?

What PMs were most important in establishing the rankings of the alternatives? How should those rankings be interpreted given the following statement?

“There remains considerable uncertainty in all of these parameters. Should the priors formulated for them poorly represent the true values, the PMs computed for the EIS alternatives could deviate considerably from what they should be and even the actual rankings of the EIS alternatives in terms of the PMs could be quite different from results found in this report.”

The sensitivity analysis is briefly summarized. As expected, the results are sensitive to mortality probabilities given the models’ structures and designs. This is likely due to the lack of compensatory processes in the LCM.

The authors then briefly discuss extensions to the LCM. The ISAB has the following comments on these:

Moving to monthly time steps may not be useful. The largest component of mortality is in the ocean, which will not be affected by moving to monthly time steps. The rationale and

anticipated improvement to the models by moving to monthly time steps warrants an explanation.

Given the sparsity of the data, moving to a full Bayesian models is unlikely to be useful with many parameters nearly non-identifiable. Improving the current models might be more efficient and productive. The ISAB has identified several aspects of the models that could be revised or improved to better inform the EIS alternatives.

Including individual covariates such as size is unlikely to be feasible since size must then be tracked through the entire lifecycle and there is little information on the processes that depend on size while the fish is in the ocean.

Editorial

p. 154. "...was still a fairly wide band of uncertainty in key performance metrics." Uncertainty usually refers to variation in estimates. As noted previously, the range of values over the simulations are not confidence intervals (i.e., no uncertainty) but variability in life-cycle trajectories.

Review of Appendix A. Spring Chinook Salmon Model Specification

This very useful appendix helps the reader determine how the model is structured.

Specific comments

Based on Table A.2, it would appear that PSM below the dam is simulated ignoring water temperature or other factors. Is this a reasonable or realistic assumption?

Age composition is assumed to be fixed over time. Is this reasonable?

Dam passage efficiency and dam passage survival are bootstrapped. It is not clear if each parameter is bootstrapped separately, or as a pair.

Table A.3. Equations do not have any demographic stochasticity. For example, the number of natural-origin spawners is a deterministic function of outplants and prespawning mortality. How are "fractions" of a fish handled? Would use of binomial distributions provide a more realistic representation and generate more realistic total variability?

Editorial

(similar comments apply to A.2.2, A.2.3, etc.)

p. 169. Use proper scientific notation rather than 6E+06.

p. 170. Table A.2 (and others) uses “parameters” in a non-statistical sense by including “constants” (such as maximum hatchery-origin outplants) that are not estimated from data. These constants should be put into a separate table from parameters that are estimated from data or assumed to be known (e.g., ocean survival probabilities) because data are not available. The use of the term “parameters” should be limited to those factors that describe the processes in the lifecycle model (those in orange boxes in Figure 2.1.1).

p. 170. How was an estimate and standard error converted to a Beta distribution, e.g., for PSM below the dams?

p. 171. What does the Triang(a,b,c) notation mean for RRS?

p. 171. What do the two Normal components for the BH egg capacity refer to? These appear to be surprisingly precise with the SD less than 5% of the mean. Are these sensible uncertainties for this parameter?

p. 172. “Harvest rates.” These are not rates (not per unit time) but rather probabilities. Change throughout the manuscript.

p. 174. Mathematical notation uses \times to indicate multiplication rather than $*$.

p. 174. It would help if the equation number (left most column in Table A.3) was overlaid on the LCM figure presented earlier, so a reader can quickly find the relevant equation for each stage in the LCM.

Review of Appendix B. Steelhead model

The model appears to be much simpler than the Chinook model with only a few estimated parameters. Similar comments to those for Appendix A apply to this model but are not repeated here for brevity.

Most parameters do not appear to vary across simulations, so uncertainty is certainly understated across the 10,000 simulations. No variation in ocean survival is modeled post 2019 (line 3 of Table B.2) – is this correct?

Review of Appendix C. CJS analysis of PIT tag data

This appendix describes the analysis of the PIT tag data using a Cormack-Jolly-Seber (CJS) model. The likelihood function (Equation C-1) makes it clear that all adult returns at Willamette Falls (WFF) are pooled. However, these adult returns are a mixture of ages and therefore do not have a common S_2 survival probability. This violates a key assumption of capture-recapture models of homogenous parameters for all fish in a cohort. So, what does S_2 then measure? This value is used subsequently in other chapters but has no meaning. In Chapter 2 (p. 27), they explained that they assume returns are predominantly age 4. The authors stated that the BCJS estimates are inadequate to use directly in the LCM and adjusted for age in the survivals used for the LCM (described in App. D). If the age structure is sufficiently concentrated at a single age, this problem may not greatly affect the survival estimates, but there is no analysis reported to support this assumption.

The ISAB is concerned that the analysis may not be correct. For example, a Bayesian analysis does not have an “objective function.” The authors describe a random walk method to obtain the posterior distribution, but the likelihood function for this CJS model is sufficiently simple that a direct implementation is easy to do. Why was this random walk Metropolis method used? The authors state that the MCMC was initialized at the MLE of the model parameters. Yet the authors acknowledge that the final survival and detectability parameters are confounded; thus, individual MLEs do not exist for these last two parameters. How was this overcome?

The authors develop priors to deal with the above issue. More details are needed. In addition, the priors, as given, appear to be highly influential (see comments in Appendix J). In cases with prior distributions with small uncertainty, a prior-posterior overlap is often computed (Gimenez et al., 2009). Checking the overlap is particularly useful when trying to determine if the parameters in the model are identifiable. If substantial overlap exists, the prior may simply be dictating the posterior distribution – the data may have little influence on the results. While a large degree of PPO is not always a bad thing (e.g., substantial prior knowledge about the system may result in very informative priors used in the model), it is important to know where data were and were not informative for parameter estimation. In this case, it would appear that the results obtained may be an artefact of the informative priors.

Editorial

p. 213. “...survival rates” “...detection rates.” These are not rates (i.e., not per unit time), but rather probabilities. Drop the term “rate” here and elsewhere in the manuscript.

p. 215. “Note that the LCM did not use from the CJS model estimation.” Something seems to be missing here.

p. 215. “The statistical likelihood is the product of the observed detection histories (i.e., 111, 110, 101, 100) and the expected probabilities for each possible detection history.” The likelihood is not a product of histories, and there are no such things as “expected probabilities.” It might improve the documentation to have a statistician review the chapter. A few examples where rewording is needed are provided below.

p. 215, “...expected detection histories” What is this? Rewording is needed.

p. 215. Equation C-1. This is not the log-likelihood (it is missing the constant terms) but is proportion to the log-likelihood.

p. 215. “... expected probability detection histories of 111,..,100.” Rewording needed.

p. 215. “...inform a prior distribution for detection site 3.” Rewording needed – the prior distribution was applied to parameters and not to sites.

p. 216. Figure C.2. The diagram uses S for survival probabilities, but Appendix J uses ϕ . The document should use symbols consistently.

p. 216. The authors tested a state-space model. Why? There are no individual covariates, and so the multinomial model is faster and equivalent to the state-space model.

p. 216. “...to ensure that the posterior distributions for the model parameters were unimodal.” Just because the posterior distributions are unimodal does not imply convergence of the MCMC algorithm. “Also, the model was tested at different initial starting values.” This implies that measures such as R -hat can be computed. Were they? These measures should be reported.

p. 216. Equation C-2. This equation does not appear to be correct. Distributions cannot simply be added to get a function to add to the objective function. In a Bayesian analysis, there is no “objective function” that is minimized. Also, the parameter P_2 is included twice.

p. 216. What is meant by the term “mean posterior estimate”? Do you mean the mean of the posterior distribution for a parameter? Rewording is needed.

p. 216. “... 0.192 (median = 0.079).” Does not match Table C.2.

p. 216. “pP2.” Something is missing.

p. 217. “...separate the SAS.” SAS should be defined.

p. 218. Figure C.3. What is “R”? Why is $R=0.83$ bolded? The shape of the joint posterior distribution of S1 and P2 is worrisome because it indicates near non-identifiability. Explanation is needed. This near non-identifiability implies samples from the joint posterior distribution for these parameters will have extreme variation, but their product will remain roughly constant. Why not sample directly from the posterior from the product?

p. 219. Table C.2. The 90% credible intervals are presented. Why 90% rather than 95%?

Review of Appendix D. Reparameterization of the CJS Survival

This appendix describes how initial values were obtained for the first-year-at-sea mortality and proportion maturing at age for all basins combined. These values serve as initial values for a fitting process where basin-specific values were obtained.

This chapter would be easier to follow if a series of equations was presented rather than text descriptions. This appendix is a list of steps with no explanation or rationale of why the steps should work. A single (Bayesian) model incorporating all of the various parts of this analysis in this chapter should be constructed rather than a series of models that use the estimates from the previous model in a chain. In this way, the uncertainty in each of the steps is properly propagated to the final results.

In Section D.3 (p. 224) an “objective function” is constructed. Why does the objective function include the value of 1000? The least squares fit does not account for the fact that the G s must sum to 1, i.e., are not independent. Why does the fit use $\log(G)$ rather than simply G ? This appears to be an ad hoc fitting procedure with little statistical validity. No estimates of uncertainty are provided or seemingly used in the LCM.

The work in this chapter is difficult to follow, and it is difficult to verify if the estimating equations have been implemented properly.

Editorial

p. 225. Table D.5. There is misalignment of rows in tables.

Review of Appendix E. Time series of steelhead marine survival probabilities

This appendix obtains estimates of marine survival for steelhead. The actual LCM uses the mean survival and yearly deviates in the fitting process. No measures of uncertainty are obtained and it is not clear how uncertainty in these survival proportions is introduced into the LCM.

Review of Appendix F. Modelling Lag 1 Autocorrelation

This chapter attempts to adjust survival probabilities for steelhead for autocorrelation.

Equation F-1. No explanation of the form of the prior distributions is provided. These look like they are based on a (log) Normal distribution but this is not clear. The survival probability is divided into an “average” and yearly deviates from the average, but the combined term does not properly represent this division. There are many “prior variances,” but these have no explanation. Why was σ_{ds} assigned the value of 1?

Equation F-2. This presents a “likelihood” function, but no justification is presented for this. Also note that technically, this appears to be the negative of the log-likelihood function.

How are these two equations used to obtain estimates? Was this a “penalized likelihood approach”? Are the “priors” of Equation F-1 included, and so is this a Bayesian analysis? How were the yearly deviates obtained? How are the deviates of Figure F.2 interpreted since survival probabilities must be between 0 and 1, but some of the deviates are quite large? Are these additive or multiplicative and on what scale? How were the values in Figure F.3 obtained? Figure F.3 should show a similar pattern to Figure F.2, but it is not additive?

No justification is given for Equations F-3 to F-5. Is this a simple AR(1) model? If so, then it should be presented as such.

Editorial

The term “prior mean value” is used throughout this chapter and in the report. The ISAB recommends proper use of the term (i.e., “mean of the prior distribution”).

Review of Appendix G. Prior for initial steelhead spawner abundance

This chapter examines a potential “correlation” between estimates:

“was large negative correlation between the estimates of freshwater survival, ϕ_F , and the estimate of initial abundance of female spawners in 1991.”

This is not surprising and is a sampling correlation. A large negative correlation indicates that these parameters are essentially confounded and the data are unable to separate them. This can lead to situations where “Some of the higher initial abundances with similar likelihood were beyond the historical range of the Foster counts, and so were not deemed credible.” The authors attempt to develop a prior distribution for the initial abundance to correct this problem.

p. 234. A log-normal distribution was used as a prior distribution for the female spawner abundance at Foster in 1991 and fit using a method of moments (using the median and CV) to give the parameter values for the prior distribution shown in Figure G.2. This is a reasonable approach.

Review of Appendix H. Development of Bayesian priors

In this chapter, the empirical prior distributions used elsewhere are derived. The documentation needs to provide greater detail and more thorough description of what the resulting priors are supposed to describe.

There were three release years for which the survival probability between Minto to Portland could be computed (bottom of p. 235), presumably from a CJS model. Each estimate would be accompanied by a measure of uncertainty (the SE) which was then used to compute a 95% confidence interval. A beta distribution was found for each year that matches the confidence limits with its 2.5th and 97.5th percentile. Then these three beta distributions are combined, but no details are provided on how this is done. Figure H.1 shows the combined prior distribution, but it is not clear what this combined prior distribution describes? Is it supposed to describe the year-to-year variation in survival probabilities? Is it supposed to describe the uncertainty in the mean (over all years) of the survival probability? More details are needed.

Unfortunately, this is not the best way to combine multiple estimates. What is better is to fit a hierarchical model where the yearly estimates are allowed to vary around a common mean, with uncertainty. There are two stages of variation in this hierarchical model: year-to-year variation in the survival probability and uncertainty around the yearly estimate.

A different method was used to obtain a prior distribution for the adult survival from Portland upstream (Figure H.2). There are multiple estimates, and a beta distribution is fit to the empirical distribution of the estimates. However, each estimate also has an uncertainty and this has been ignored. Again, a hierarchical model as described above is the proper way to combine the estimates.

A two-step method was used to determine a prior distribution for the detection probability at SUJ. First, a series of release cohorts was used to estimate route-selection probabilities and detection probabilities based on radio- and PIT-tagged studies (Table H.1). Then a proposed prior distribution is obtained by sampling using a uniform distribution in the range of each component from the multiple release groups. Again, a hierarchical model described above is a superior way to do this. This “mechanistic” prior distribution does not account for differences in flow, and so the second step is an adjustment for flow using a hierarchical model. This is a reasonable approach, except that the “sigma” parameter does not appear to be computed properly (see notes below). Then a beta distribution is fit for each flow level by matching the mean and variance from the posterior distribution from this hierarchical model.

A Beta distribution was chosen for the detection probability at WFF. How were the values of 191 and 3.9 for the parameter value chosen? These do not match the values used in Appendix J. Why?

Editorial

p. 235. The text states “models the survival rate between discrete release and detection locations by adjusting the numbers detected by the detection probability at each location.” The intended meaning is difficult to discern and could be improved by rewording.

p. 235. The text states “mean and 95% confidence interval (CI) for the survival probability of each release from Minto to Portland.” Each release will have a single estimated survival probability. So, what is the mean computed over? Do the authors apply that mean of the survival estimates over the releases? If so, then what does the confidence interval apply to? How was it computed?

p. 236. “...using the beta.parms.from.quantiles function. This function is not available in Base R. Which package or source does it come from?”

p. 236. “...a beta distribution was fitted to all three empirical distributions combined to.” How was this done?

p. 240. “logit(σ).” Sigma is not a probability, so a logit function cannot be applied to it. This is not just a typo because this also appears in the R code in H.3.4.

p. 241. Equation H.2 is on the logit scale, but sigma is computed assuming variability on the [0,1] scale. These are not compatible.

p. 241. Figure H.4. Better choice of colors for gray/blue. Is there only 1 medium flow data value?

Review of Appendix I. Bayesian prior to account for tag loss etc.

Adjustment factors were estimated to account for effects of tagging mortality, tag loss, and differential survival based on fish origin. Point estimates were obtained from the literature and the (see comments on Appendix C and J) apparent survival probabilities from the CJS model were simply multiplied by these factors (Equation I-2).

A simulated prior distribution for the adjustment factor is created by a simple simulation study (Figure I-2). Then, as noted by the authors

“To account for these three factors, we formulated a prior distribution for an adjustment factor for tagging-based estimates of survival rates for different salmon life stages that incorporated plausible ranges of values for tagging mortality, tag loss and differential survival based on fish origin.”

The authors then state

“We did not include this prior in the parameter estimation. Instead, the apparent survival estimates were adjusted after parameter estimation.”

The authors provide no details on how this is done.

The main body of the report (p. 16) also states

“The adjustments for tag-induced mortality rates, tag loss rate and hatchery effect were applied to the posteriors for survival rates from the Bayesian CJS methodology to formulate adjusted posterior distributions for downstream juvenile and total smolt-adult survival rates.”

but no details are provided here either. Again, note that total smolt-adult survival probability has not been estimated properly (see our comments on Appendix C, D, and J).

Similarly, the main report (p.24) also states this adjustment procedure is done but no details are provided on how this is done.

ISAB Review of Appendix J. Simulation-estimation analysis of Bayesian Cormack-Jolly-Seber model

The ISAB reviewed Appendix J in the UBC report, and the UBC modelers also provided additional analyses and discussion in response to questions about the appendix from the ISAB.

Given the range in release sizes and low numbers of adult returns to Willamette Falls observed in Upper Willamette PIT tagging studies, this UBC study attempts to examine the performance estimates of survival under a range of release sizes and detection histories from a Bayesian capture-recapture analysis. In particular, the authors are exploring the effect of having no adult fish detected after release.

The information in Appendix J would be informative for exploring the behavior of the estimators. Many other studies simply apply the statistical method to the data without a simulation study to assess the performance of the estimators. While Appendix J is not used directly in the life cycle model, its conclusions are used to justify decisions on other parts of the report (e.g., Appendix C). Consequently, UBC modelers could consider the ISAB's comments and suggestions in future updates of the UBC model documents.

Definition and analysis of bias

There are two potential sources of bias: procedural and statistical. Procedural bias occurs because of flaws in the experimental or measurement processes – non-random sampling, error in instruments, errors in data processing, and such. Procedural biases can be reduced to acceptable limits through careful experimental design and data collection methods and is assumed negligible here.

Statistical bias occurs when the statistical method used to estimate a parameter may not reliably and consistently estimate the parameter of interest because the methodology is incorrect (e.g., assuming a normal distribution for the data values when this is not appropriate) or is a feature of the estimator (e.g., estimates of ratios can be biased even if the numerator and denominator are unbiased).

Statistical bias is quantified as follows. An estimator attempts to measure a parameter (θ) by drawing a probability sample and computing an estimate ($\hat{\theta}$). It is unlikely that the estimate will equal the true parameter value – in some cases, it will be higher than the parameter value and, in some cases, lower than the parameter value. For example, if you flip a fair coin 3 times and measure the proportion of heads over the 3 flips, the possible values are 0/3, 1/3, 2/3, and 3/3, none of which is equal to the (assumed) true parameter value of 0.5. The distribution of the estimates over all possible samples is called the sampling distribution.

In many situations, the sampling distribution cannot be evaluated theoretically. Rather a simulation study is performed where multiple sets of simulated data are generated (a total of n simulations), and the estimate is computed for each simulated data set. Let $(\hat{\theta}_i)$ represent the estimate from the i^{th} simulated data set.

The estimated statistical bias is then found as

$$\widehat{Bias} = \bar{\hat{\theta}} - \theta$$

where $\bar{\hat{\theta}} = \frac{\sum \hat{\theta}_i}{n}$, i.e., the arithmetic mean of the estimates over the simulations. Again, the estimated bias is a single number and is not attached to any particular data set.

In Appendix J, Equation J-4 was used to compute a “percent relative bias” for each data set. This is more properly defined as the relative deviation of an estimate from the true value. Consequently, we suggest that discussions of the distribution of relative bias (Figure J.1), range of bias, or interquartile range of bias should be revised because bias is a single number. The computed statistics are more properly designated as properties of the (relative) sampling distribution. Similarly, statistics such as median bias or mean bias (Table J.1) are not defined because bias is a single number. Bias cannot be associated with a particular dataset. Based on this perspective about bias, we suggest that most of the discussion in the appendix about bias should be revised. We do agree that the mean relative deviation over the sampling distribution is the relative bias.

The ISAB asked the UBC modelers about the analysis of bias, and they agreed that what they discussed in Appendix J is not bias, but rather the distribution of percent relative error of posterior mean estimates from true values. They noted that the values in Figure J.1 would not change but should be labeled as distribution of percent relative error of posterior mean estimates from true values.

The ISAB appreciates the corrections and thorough discussion of bias and its effect on the analysis. We encourage the UBC authors to include their response’s discussion of bias and its potential consequences for the Corps decisions about EIS alternatives. The UBC analysis shows that substantial relative bias could exist in the estimates. We agree with the authors that the impact of bias may be attenuated when model calibration is performed. We also suggest that the author move from separate individual analyses of individual release groups to a combined analysis of all releases groups with a hierarchical structure on the yearly parameters. This should alleviate many of the concerns expressed about convergence of individual release group fits. The ISAB makes a similar recommendation about the methods in Appendix C. We also

recommend that the modelers include specific conclusions in revised versions of Appendix J to help the reader understand the figures and tables.

Some of the problems caused by a bias in one part of the model are “corrected” when a calibration against known returns is done. However, the amount of data available for calibration is limited and so the potential for bias correction is also limited.

The ISAB agrees that if some bias exists when considering the alternatives, relative comparisons presumably are still useful, but absolute performance may not be realized in the real world. The real concern is dealing with P(QET) since positive biases in abundance from the model could dramatically lower the reported P(QET).

The ISAB would also like to note that the input from the FBW might substantially influence model outcomes among the alternatives. Therefore, bias in the downstream portions may have only a slight influence on the relative performance of the alternatives.

Questions about adequacy of numbers of simulations

Normally, one conducts many thousands of simulations so that the uncertainty in the estimated bias is very small. The authors only did 100 simulations because it was too computationally intensive. This is an artefact of the very slow method to fit the Bayesian CJS model, the state-space model. The state-space model should only be used if there are individual covariates or more complex processes. In this case, the authors could have created a Bayesian model using a multinomial likelihood with associated prior distributions on the parameters. This can be implemented runs done in fractions of seconds even for very large sample sizes. For example, the following source⁵ gives the BUGS/JAGS code to implement both the state-space and Bayesian multinomial models (for CJS model with many more capture times). In the case of 2 or 3 capture times, the "mij" array is considerably simpler. The authors concluded that the analysis can be speeded up greatly by aggregating the capture histories into an m-array and they get the same results as the state space parameterization provides.

Additionally, with a small number of simulations, a confidence interval for bias should be computed around the estimated bias to ensure that small apparent biases are not just sampling artefacts.

The ISAB asked the UBC modelers about the adequacy of numbers of simulations. The UBC modelers agreed with our suggestion to explore the Bayesian m-array formulation of the multinomial model in JAGS. It allowed them to explore a much wider set of simulations while

⁵ <https://bcss.org.my/tut/open-population-models/cjs-models-for-apparent-survival/>
<https://bcss.org.my/tut/open-population-models/cjs-models-with-aggregated-data/>

conducting 1,000 simulations per set. The UBC modelers then examined the performance of the CJS model in estimating each parameter given a range of “true” values in an updated simulation study. The UBC authors found that the vague uniform (0,1) prior tends to pull estimates of parameter that are <0.5 towards 0.5 and estimates of parameter that are >0.5 towards 0.5 (i.e., pulls values towards 0.5 from either end). In particular, the ISAB noticed the changes in the estimated means as compared to the true values of the parameters in their analysis of informative priors. We agree that the informative priors thus appear to be most useful when the release sizes are low, for example, in studies that use beach seine sampling.

While some of the relative biases presented in the UBC’s response appear to be large, part of the reason for these large relative biases is the very low values for the SAS. Therefore, a difference of 0.01 when the base probability is 0.01 is a 100% relative bias, but if the base probability is 0.10, this is only a 10% relative bias. Consequently, the figures provided in the response are more helpful in determining the absolute size of the potential bias. We encourage the UBC authors to include this information and figures in future documentation of the model.

The additional analyses the UBC modelers conducted to respond to our questions are useful. Future revisions of Appendix J would be improved if summary conclusions were also presented about bias. As noted previously, some of the bias may be attenuated during the calibration phase, and, if biases are consistent across the alternatives, relative comparisons may still be useful.

Assumptions of homogeneous survival between SUJ and WFF

The simulation assumed a constant survival for all fish between the Sullivan Dam Juvenile Bypass facility and detection at the Willamette Falls Adult Fishway (ϕ_{WFF}). However, the adult fish that return to the WFF are a mixture of multiple ages and therefore do not have a homogeneous survival probability (some fish have 3 years of ocean survival; some fish have 4 years of ocean survival, etc.). The constant survival assumption used in the simulation should be revisited. The Bayesian Cormack Jolly-Seber model (CJS) that was used to fit the simulated data (and also used in Appendix C) also appears to oversimplify potentially important biological processes with multiple maturation ages, harvests of different age classes, and returns that include fish from multiple release cohorts. The approach used in this chapter could be useful to assess whether the incorrect Bayesian CJS model gives sensible estimates when applied to simulated data if the model is revised to reflect more realistic assumptions on survival and ?.

The ISAB asked the UBC modelers about homogeneous survival between SUJ and WFF. The UBC group responded and examined the Chinook salmon PIT tag data from the North Santiam, McKenzie, and Middle Fork subbasins for individuals that were detected at both SUJ and WFF to determine proportions spending different numbers of years at sea. They found across all

subbasins that >90% of fish return to spawn at age-4 or age-5. The proportions varied across the subbasins, and the potential bias in the assumption of homogeneous survival is greater for the estimates for the McKenzie River subbasin. They used these relationships with the CJS model to calculate percent relative bias for simulations. The scale of the estimated bias, given an input value of the survival from smolt-adult returning at age-4, depends on both the release size and priors used. As expected, the difference between the two bias values was larger when the proportion returning at age-4 was low, and smaller when the proportion was closer to 100%. The ISAB appreciates the detailed response. The biases appear to be considerable, especially for small numbers of releases. We hope that calibration would alleviate the impact of these biases. A summary of this analysis and its findings will be useful information in revised model documentation.

Methods to assess the effect of conditioning on non-zero returns to WFF

The UBC authors are concerned that cohorts with no recoveries at WFF provide no information and wish to condition on cohorts with at least one detection at WFF. They did this by selectively looking at results for simulations that had had a least one detection at WFF. We question this approach because the analysis of each simulated dataset did not condition on having at least one detection at WFF.

As an analogy, suppose you wanted to study the number of fish caught in a sample of fishers. A sample of fishers is selected, the number of fish recorded for each fisher. You first fit a Poisson distribution to this data. A likelihood is constructed that uses the full Poisson distribution. But you are concerned that there may be 0 inflation (i.e., some fraction of fishers have very poor skill). So you wish to only use fishers who caught at least 1 fish (conditioning on positive counts). The full Poisson likelihood would not be appropriate because that assumes 0 counts are possible. Rather, the likelihood and the analysis should use a Truncated Poisson distribution to accommodate the conditioning on at least one fish captured.

In the same way, an analysis to study the impact of conditioning on at least one recovery at WFF must account for the conditioning process. A separate simulation study that generates datasets specifically with at least one recovery should be used, and then a modified CJS likelihood and Bayesian CJS model (to account for the conditioning) should be fit.

We noted earlier that 100 simulations are too few. The documentation indicates the authors proposed to drop data based on only 29 simulations, which appear too low.

It is important to note that the state-space model they use for the Bayesian CJS model cannot be modified to accommodate this conditioning. Modifying the multinomial likelihood-based Bayesian CJS is a straightforward task.

The conclusion to ignore cohorts with no recoveries at WFF is not supported by this appendix and needs to be revisited. This data filtering is not mentioned in Appendix C nor in the main report, therefore it is not clear if data sets with no adult recoveries are actually removed. This also needs further clarification in the report. Such important data filtering warrants thorough discussion in the report, in addition to an appendix.

The ISAB asked the UBC modelers about assessment of the effect of conditioning on non-zero returns to WFF. The UBC modeling team agreed that their initial number of simulations (100) was too low. A major effect of the data filtering decision was that the model did not generate estimates from any releases of PIT-tagged fish in the South Santiam because no tagged fish were detected at WFF. They also did not obtain estimates from all years of natural-origin releases in Middle Fork and North Santiam rivers because release sizes were typically <1,000 fish and few returning adults were detected at WFF. The UBC authors indicated they will revisit their decision and model all release groups.

The ISAB agrees that the updated simulation-estimation results suggest that the decision to exclude data should be revisited and it would be better to model all release groups. We also suggest that rather than analyzing each release group separately, a combined model of all release groups with a hierarchical structure across the years on the SAS and detectability parameters would be even more beneficial (see our detailed comments about Appendix C). Such a combined model is not that difficult to code in BUGS when the multinomial likelihood approach is used and would alleviate many of the concerns about convergence.

The revised Appendix J should clarify if the decision to drop data points was changed.

The UBC response also noted that the LCM might accurately predict the historical data and could be used to accurately rank relative performances of different policy options even if some survival rate parameters might be too high or too low. An individual parameter could be biased, but other model components could have compensating biases, and the fitted model could still accurately predict responses and accurately rank the performances of alternatives. We agree with the UBC comment about “compensatory” changes in parameter estimates when data is calibrated. The lack of a long-time series for calibration is unfortunate.

Effect of extremely informative priors

The prior distributions used in the Bayesian CJS are so informative (see below) that a properly conducted simulation would likely just reflect the informative prior and not the influence of the data.

For example, the prior distribution for ϕ_{SUJ} is a Beta(2.440, 3.665). One of the prior distributions has an alpha value < 1. This can lead to odd shapes of prior distributions. The prior distribution for ϕ_{WFF} is very informative as a Beta (3.193, 277.76). For example, Figure 7 presents plots of the prior distributions along with the “true” parameter values used in the simulation:

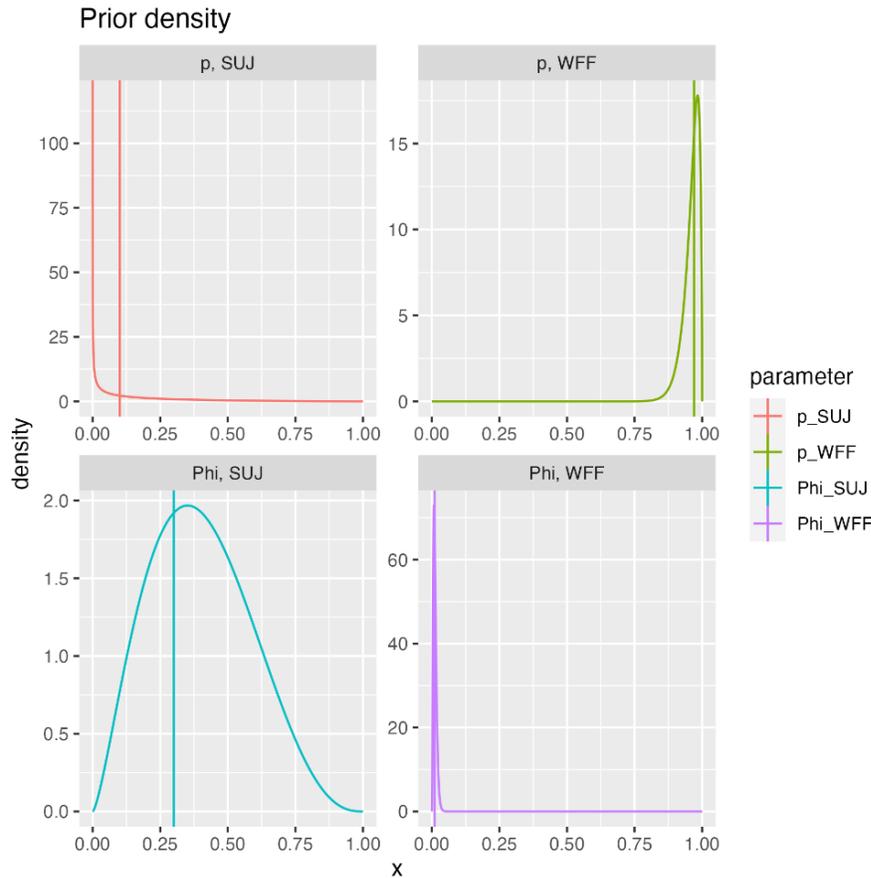


Figure 7. Prior distributions versus the “true” parameter for the ϕ_{SUJ} simulation in Appendix J.

The prior distribution for detectability at SUJ has a very odd shape. Is that correct or appropriate for the analysis? This prior distribution has most of its mass well below the simulated parameter value, which by itself likely causes bias in the estimates. The prior distributions for the WFF site are very informative and likely overwhelm the data. The prior distributions seem to have been determined in Appendix C, but Appendix C also does not address the effect of this very informative prior distribution.

In cases like this, a prior-posterior overlap (PPO) is often computed (Gimenez et al. 2009). Checking the overlap has particular utility when trying to determine if the parameters in the model are identifiable. If substantial overlap exists, the prior distribution may simply be dictating the posterior distribution – the data may have little influence on the results. While a

large degree of PPO is not always a bad thing (e.g., substantial prior knowledge about the system may result in very informative prior distributions used in the model), it is important to know where data are and are not informative for parameter estimation. In this case, the results obtained may be an artefact of the informative prior distributions. Appendix C should also contain information about the effect of the informative prior distributions on the estimates.

The authors present the “mean PRB” (but no confidence interval), which we think is incorrect, and do not discuss if these results are simple artefacts of the informative prior distributions. For example, the apparent positive bias in the estimates of ϕ_{SUJ} and negative bias of ϕ_{WFF} would appear to be artefacts of the prior distributions (see above plots). The results are not intuitive and require explanation. As the number of releases increase, the effect of the prior distributions should diminish and the estimates should be closer to the MLE, which usually implies that estimates should be unbiased. Yet Table J.1 shows absolute bias increasing with large sample sizes. It is not clear why this occurs. This may be an artefact that with large sample sizes, the prior distribution becomes less and less relevant, but confounding of survival and catchability at the final site becomes more of a problem. More detailed explanation of these results would provide useful information for understanding the analysis.

The ISAB asked the UBC modelers about the effect of extremely informative priors. The UBC response indicate that they agree that the prior for pWFF (detection probability at WFF) is highly informative. They explained that they wanted to estimate SAS (smolt-to-adult survival rate) separately. Detection probability at fish ladders in the Columbia is very high, and use of highly informative priors provides some uncertainty in the parameter. They agreed that priors with CV much lower than 0.5 are highly informative and may overwhelm the data; however, the updated simulation-estimation analyses found that the data are not overwhelmed by these priors, even when the true values used are towards the tails of the prior distributions. They considered the shaped of the distributions at SUJ to be valid and indicated they do not understand the ISAB’s comment.

The ISAB found the UBC response to be very useful for clarifying the process by which these priors were chosen. We note that Figure 5. in the UBC response has two identical panels in the graph, which might need to be corrected in future reports of the analysis.

Editorial comments

p. 252. Why was the median of the posterior used to study bias? The mean of the posterior would seem to be the standard choice.

p. 253. “Accurate” has a specific statistical meaning that is not computed here.

C. Review of the NOAA Models

1. Summary Comments

The NOAA life cycle models (LCMs) were developed as an extension to an earlier LCM (Zabel et al. 2015) and reviewed by the ISAB (ISAB 2014-4). A document summarizing the results was provided to the ISAB, along with a briefing to the ISAB followed by a question-and-answer session. After a preliminary review, the authors provided responses to a set of questions from the ISAB in a presentation and a written document.

The NOAA LCMs follow a hybrid approach. The sub-components of the models (e.g., egg deposition, fry survival, dam passage efficiency, etc.) are developed as “standalone” components and, except for a small amount of calibration, no overall model fit to estimate all parameters in the LCMs is done.

Model parameters were derived from studies in the basin, borrowed from nearby populations, derived from the model calibration process, taken from published studies, or derived from expert opinion. The FBW provided key inputs for the impacts of the alternatives on dam passage and survival and the FBW is treated as a “black box.” Calibration was to a limited set of data by adjusting (a small number of) parameter values so that predictions matched the current data.

The LCMs were then used to simulate 100 years forward using historical water years. No modeling of potential climate change effects on the alternatives was performed. Consequently, comparisons based on historical water-years may not reflect performance of the alternatives in the future and some care is needed in interpreting the results.

Viable salmonid population (VSP) scores were created to summarize abundance and productivity (primarily through the probability of quasi-extinction), hatchery contributions, and life-history diversity with most weight given to the abundance and productivity VSP score that is summarized by the P(QET). There are a number of issues regarding how the report estimates P(QET) and then converts it to a VSP score.⁶

Uncertainty was included in many (but not all parameter values) but was not included with the outputs of the FBW that potentially dominant the LCM responses to the different water management alternatives. Consequently, the predicted quasi-extinction risk and the derived VSP scores may be biased.

⁶ Note that the VSP scores were not ultimately used in the multi-model analysis.

If the primary driver of the differences among the alternatives is the FBW inputs, then it is not clear how much value added is provided by the LCMs. In this situation, the LCMs acts as complicated bookkeeping that express the FBW results in terms of VSP scores and the internal dynamics of the LCMs become minor influencers. Determining the influence of the FBW and the other supporting models to the LCM predictions versus how the internal dynamics of the LCMs affect predicted responses is critical to proper interpretation of the results.

The document places a heavy reliance on the total VSP index which is a summary of outputs. The figures in the Summary seem to show no clear alternative performs best and the perhaps alternatives 1 and 4 have promise, but the degree of responses in the components (e.g., abundance versus diversity) of the VSP index is not discussed. Later in the report, the components of VSP are reported, but additional discussion is needed. Graphs of actual model summaries may also be helpful, e.g., show the distribution of the P(QET) directly.

The ISAB review below consists of summary responses to a set of questions developed for this review and a detailed review of the NOAA-LCM report.

2. ISAB Answers to Review Questions

The ISAB has developed the following sets of questions for the review based on questions provided by the Corps and its internal discussions.

1. Model structure and development

- a. *Is the model structure appropriate for evaluating fish responses for the EIS alternatives?*
 - i. *Are the relevant processes (mortality, growth, reproduction, movement) included and are they represented in a way that encompasses the changes within and among the alternatives?*

The NOAA models include, by definition, the major processes involved in a population trajectory. The models seem reasonable. The dependence of the results about projected impacts of the alternatives on the supporting models (especially the FBW) needs to be assessed. Because the FBW and other supporting models are treated as a “black box” and simply used with the LCMs, any structural problems or uncertainties within the FBW and other models will be transferred and propagated through the LCMs.

- ii. *Are the key functional forms (e.g., shapes, dependence on hydrological and environmental variables) sufficiently realistic?*

In general, the functional forms (submodels) are well reasoned; the ISAB identifies a few general comments and specific comments for each component as noted in our “Detailed Comments” section (below).

iii. What aspects of the model are stochastic?

The stages in the LCMs appear to include a reasonable representation of stochasticity, with the exception of the variability of the outputs for the FBW and other supporting models. Consequently, the variability in the LCM responses is likely too small.

b. Are the models formulated to assess fish responses at the appropriate temporal and spatial scales that both capture the effects of the alternatives and generate outcomes relevant to management scales?

The temporal scale is annual over a 100-year period. The FBW results on dam passage and survival are available at finer temporal scales and are combined to produce a single value for each year. This is a sensible scale for the LCMs developed here but could miss important responses to within-year variability because the response to an annual mean condition is not necessarily the same as the mean response to, say, daily conditions. The averaged conditions used as input to the LCMs should be compared with the finer time scale results to assess whether additional stochasticity (e.g., extremes) also should be examined with the LCMs.

c. Are the estimates used for model input parameters (hydrological, environmental, processes) reasonable and scientifically defensible? What data are used to estimate or confirm model parameter values?

Model parameters were derived from studies in the basin, borrowed from nearby populations, taken from published studies, or derived from expert opinion. Finally, some calibration to a (small) set of current data was used to inform some parameters. The report should summarize all the parameters used in the LCMs with a summary table indicating the source of the parameter values to aid the reader in understanding their values.

d. What sources of variation that affect population dynamics and responses have been accounted for?

Variability in environmental variables is modeled by using the set of historical water years and starting at a random point in the series. Uncertainty in parameter values is accounted for by sampling from a distribution centered about the population estimate with suitable variability. However, some parameter values are fixed (e.g., ocean survival, which is highly influential and variable). Population stochasticity between LCM stages is modeled using appropriate statistical

distributions. The FBW output for each alternative does not incorporate any stochasticity. Consequently, overall variation in population dynamics may be underestimated.

- e. Have the researchers adequately evaluated the sensitivity of the predictions to uncertainties in various sources of data?*

There is no discussion of sensitivity to data in the report. However, the 2015 Life Cycle model (Zabel et al. 2015) reported a sensitivity analysis, so some prior information is available. The report should consider the earlier results and discuss if the new model has similar sensitivities.

- f. How sensitive are predictions to uncertainties in hydrological conditions, environmental variation, and model parameters?*

There is no analysis of sensitivity to these factors presented in the report. The final results appear to depend on how the FBW predicts changes in dam passage and survival. As noted in our detailed comments, the FBW is treated as a “black box” input to these models. It does not include any stochasticity and uncertainty is not included with its outputs used by the LCMs.

Potential effects of climate change on the alternatives were not modeled. Consequently, comparisons based on historical water-years may not reflect performance of the alternatives in the future, thus care is needed in interpreting the results.

- g. Are subcomponent models adequately integrated into the larger model?*

The NOAA LCMs integrate all the life-history components into a single model that produces forward predictions. This appears to be relatively seamless.

2. Key assumptions and limitations

- a. Are the model assumptions and limitations sufficiently documented? What are the key assumptions, strengths, and limitations?*

The assumptions for each part of the LCM are described in each chapter and are not listed in a summary table. It would be helpful to have a summary table of the key assumptions made for each component of the LCM. The FBW is treated as a “black box” and the reality of its output or internal assumptions are not assessed. Because the differences in the FBW output under each alternative may be important drivers of the differences in the LCM VSP scores, the internal assumptions and structure of the FBW are crucial. The FBW has been reviewed by the ISAB elsewhere ([ISAB 2014-3](#)) and is briefly reviewed above. Additional analyses of model results and new simulations designed to explicitly determine the role of the FBW and other supporting models versus the internal calculations of the LCMs are warranted.

- b. Is the reasoning, rationale, and evidence for the key simplifying assumptions (e.g., no interactions with other species) reasonable and adequately documented?*

This is not discussed in the document.

3. Model validation and verification

- a. Does the model documentation adequately describe testing steps utilized during model development (i.e., consistency check, sensitivity analyses, calibration, validation)?*

The document is silent on these issues for each sub-component, but there is limited calibration done to a (small) data set to modify a few parameters to get reasonable values for the baseline case.

- b. Have the researchers adequately assessed the fit of model to current data? Have the researchers assessed the fit of the model to new data (e.g., use it to predict years not used in fitting and compare)?*

Unfortunately, only limited data, especially the different hydrological conditions and long-term records of the abundance of juvenile and adult Chinook salmon and steelhead, were available to calibrate the LCMs. Relative results among alternatives are more robust than the absolute results (NOR spawners, recruits/spawner, SAR, pHOS, P(QET), mean QET).

- c. Has the model programming or system components been tested for computational correctness and associated errors? If not, what is the potential for errors to occur?*

This is not discussed in this report, nor in Zabel et al. (2015). In any complex life-cycle model, errors in subcomponents are propagated progressively through the model, so large errors in results can stem from small errors in multiple subcomponents. The authors should provide details on how component-wise testing of each part of the life-cycle model was performed.

- d. Has sensitivity and uncertainty analyses been applied to the model and the results used effectively to identify sensitive parameters and quantify variability around model predictions?*

This report does not contain any information on a sensitivity analysis. However, the 2015 Life Cycle model report (Zabel et al. 2015) did perform a sensitivity analysis, so some prior information is available. The report should consider the earlier results and discuss if the new models have similar sensitivities.

4. Data gaps

- a. *For which part of the model is data strongest and weakest? Is the model sensitive to these data gaps?*

Data gaps were not discussed in the report, except for a small discussion on page 137. In general, the lack of steelhead data specific to the Willamette basin leads to relatively high uncertainty in the WVS-specific predictions for steelhead. Obtaining site-specific information for the critical inputs used to develop parameters would reduce model uncertainty.

- b. *If so, then what type of research program is needed?*

Given the need to better represent natural variability and improve the data available for calibration, the Corps could support research to obtain critical data for the more important parameters in the models. Development of life history parameters that are based more from the Willamette River basin would provide better inputs for this and other models of Willamette salmonids.

5. Model output

- a. *Can the outputs of the model be used to rigorously compare and contrast different management actions?*

The LCM results were converted to VSP scores for final comparison among alternatives. The relative results among alternatives should be relatively straightforward, and comparisons of absolute predictions should be used carefully and cautiously. As noted in our detailed comments, results are potentially highly dependent on the predictions by the FBW of the differences in dam passage and efficiency among the alternatives.

- i. *How reliable are the comparisons of the effects of the alternatives (e.g., can be used only to rank alternatives; can be used in cost-benefit analyses)?*

If a key driver of the LCM results among management alternatives are the differences in dam passage efficiency and survival from the FBW (or from one of the other supporting models), then this would shift the assessment of realism from the LCMs to the realism of the supporting model. The ISAB did not review the supporting models, except for a brief review of the FBW. In addition, important forms of stochasticity may have been ignored in the LCMs that could affect the interpretation of differences among alternatives. Consequently, the results as both rankings and in absolute terms should be carefully interpreted. Very similar VSP scores may be ordered via ranking but yet essentially are the same from a biological standpoint. Absolute predictions

(e.g., abundances) have higher uncertainty than relative predictions (change in abundance between two simulations).

ii. Is the output stochastic so questions about quasi-extinction can be answered?

The output is stochastic, but the possibility of important sources being ignored (e.g., FBW passage effects) and the lack of strong calibration to past data means that inferences about quasi-extinction should account for potential biases and uncertainty. Also, the quasi-extinction probabilities are translated into VSP scores and not compared directly. Relative comparisons among alternatives should be possible. There may also be issues on how P(QET) is estimated as noted in our detailed review.

iii. Do the researchers distinguish whether differences in predicted responses among alternatives are biologically meaningful?

No, VSP scores do not lend themselves directly to such comparisons.⁷

b. What specific metrics or types of output are provided to the user? Are these outputs sufficient for evaluating the EIS alternatives? Would additional types of output be useful?

The report uses VSP scores to compare the alternatives. In some cases, it may also be helpful to view the underlying metrics that make up the VSP scores to understand the differences in the VSP scores. Some figures are presented on the underlying metrics, but the discussion of these needs improvement.

c. How far outside of the envelope of data used to develop the model will climate change take us? Do we have good information on how the model will perform outside of its development based on historical conditions?

No information is presented on impacts of climate change or model performance outside the range of data for which the models were calibrated.

6. Model improvement

a. What improvements can be made to address any critical shortcomings of the models?

A main shortcoming of the models is the lack of full propagation of input errors to the output, especially regarding the use of the deterministic FBW as a key input. This could be improved by better consideration of variability in passage survivals and other key parameters. Also, the

⁷ Note that the Corps used the component variables of the VSP separately in the multiple model analysis.

performance of the models under climate change is unknown and could be addressed by incorporating climate variation more directly into model inputs.

b. Where should future model development be focused?

The most important improvements would not come from modifying the models *per se*, but by improving the model inputs to better represent natural variability and improving the data available for calibration.

3. Recommended Changes throughout the Document

The manuscript uses the term "confidence intervals" to describe the uncertainty in the VSP score, but these appear to be percentile ranges. For example, compare Figure 1.8 and Figure 12.5.5 under the NAA. Figure 1.8 shows a range of VSP for abundance and productivity with a median of about 2.4, which matches Figure 12.5.5. The RANGE of VSP scores runs from just over 0 to just under 4 which matches Figure 12.5.5 if the error bars are 95% RANGES (i.e., the 2.5th percentile to the 97.5th percentiles) rather than confidence intervals. This was confirmed in our meeting with the NOAA team and in their response to questions. This needs to be corrected throughout the document.

Figures that compare the percentile ranges use the median to represent the center of the distribution and whiskers to represent the 2.5 to 97.5th percentile range. However, it is not possible to tell from this display if the VSP scores are concentrated around the median or if the VSP scores are more uniformly distributed. Violin plots or box plots (with modified whisker ranges) may be more useful here.

4. ISAB Comments on NOAA Report Chapters

Review of Executive Summary

This is a high-level summary of the findings of the report.

The key differences among the alternatives that affect the LCMs are those parameters that describe fish passage and survival over the dams (as predicted by the FBW), and water quality conditions (temperature and TDG) below the dams. Other factors that could influence survival rates in the life cycle models were the same, or nearly so, across NAA and the alternatives.

While these other factors would likely not change the relative rankings of the alternatives, these other factors do provide additional sources of variability, which, for example, affects the probability of falling below the quasi-extinction threshold.

The model comparisons used historical water-years when deriving the simulation trajectories. Potential effects of climate change on the alternatives were not modeled. Consequently, comparisons based on historical water-years may not reflect performance of the alternatives in the future and results should be interpreted carefully.

The plots comparing the VSP scores could be improved as noted earlier.

The variability in VSP scores for several alternatives is quite small, and it is not clear if this is because the models fail to account for all sources of stochasticity or if the VSP metric is insensitive under certain conditions. This needs to be discussed.

Review of Chapter 1. Introduction

This chapter starts with a high-level overview of the LCMs (Figure 1.2). Model parameters were derived from studies in the basin, borrowed from nearby populations, derived from the model calibration process, taken from published studies, or derived from expert opinion. Subsequent chapters describe the sub-models in more detail, but there is no overall summary of the model parameters and their sources (similar to the OSU model supplemental material). While fine detail may not be needed, it would be helpful to know what the driving variables are (e.g., dam passage survival, or water temperature, or TDG, etc.) for each part of the LCM models and the form of the relationship (e.g., a logistic function for survival). If parameter values were modified by the calibration process, there does not appear to be a summary of the original and final values (after calibration).

The importance (or lack thereof) of some simplifying assumptions could be discussed more thoroughly. For example, the models assume that no straying occurs. In the Columbia River basin, straying by fall Chinook salmon can be considerable, but tends to be lower for spring Chinook (Westley et al. 2013). Post-spawning mortality (PSM) is estimated to be affected by temperature, which is reasonable, and climate is projected to affect temperature and flow timing; thus, the importance of climate estimates and variation likely are considerable. The complex ways in which climate projections affect outputs should be noted.

A simulation approach is used to model a population of salmon or steelhead over 100 years. Parameter values are selected from a range of plausible values that accounts for uncertainty in the parameters (e.g., a normal distribution centered on the estimate). Not all parameters are allowed to vary (e.g., ocean survival is fixed at 0.8), and parameter values from the FBW and other supporting models under the alternatives are fixed. Again, a summary table of which

parameters include uncertainty would be helpful. Some parameters are unaffected by the alternatives (e.g., ocean survival), and a summary table would again be helpful.

After parameter values are established, the population is initialized, allowed to “burn-in” for five to seven years, and then run for a 100-year period using actual water-years data on precipitation and flow. A random start in the historical data is used, and the simulation then “circles” back when the end of the historical water data is reached. This ensures that “hidden correlations” in water years (e.g., ocean conditions may be related to water-year conditions) are retained. Each alternative has 1000 model runs.

For each model run, three Viable Salmon Population (VSP) scores are computed (Figure 1.3). The productive/abundance from a model run is used to estimate the probability of (quasi)-extinction (variability in P(QET) using a short cut described below which is then translated to a first VSP score. Life history diversity and pHOS are combined to give a second VSP score; finally, spatial structure gives a third VSP score.

The report describes how the model output is used to compute the VSP for P(QET), but the report was difficult to follow, especially the description of the short-cut to estimate the variability P(QET). In some cases, the description in the report does not match what was done. For example, on page 18, the authors state

“We computed P(QET) for each alternative by compiling the proportion of 1000 runs that fell below the QET threshold.”

But the summary output of the VSP scores uses the median of the P(QET) estimated using the short cut below.

Ordinarily, to estimate the variability in P(QET) for a particular set of parameter values, it would require, for example, 1000 sub-runs using the same set of parameter values (but with environmental variation in survival) because each sub-run only returns zero or one depending on whether a trajectory of the LCM fell below the QET threshold or not. The P(QET) would then simply be estimated as the proportion of the sub-runs where the QET threshold was satisfied. This direct method would require 1000 sub-runs for each of the 1000 parameter sets simulated for each alternative or a total of 1,000,000 runs and sub-runs for each alternative, which may be computationally infeasible.

The authors used a “short-cut” by only using 1 sub-run from each parameter set. This single sub-run reported either a QET event or not. Presumably, runs from parameter sets that are similar should give similar results, and P(QET) is likely related to characteristics of the trajectory of this particular sub-run. The modelers created a logistic regression model to predict the

P(QET) given the productivity and abundance from a sub-run (equation in the middle of page 20). The “productivity” for the single run was computed as the geometric mean of recruits per spawner for the first 25 years of the simulation. Similarly, “abundance” was computed as the geometric mean of the abundance in years 26 to 100 of the simulation. This logistic model was fit to the results of the 1000 sets of parameter values and the fitted model was then used to estimate the P(QET) from a single run of a parameter set which then gave the first VSP score from Figure 1.4 for each parameter set.

The authors compute productivity as recruits/spawner at "relatively low abundance" (page 18), by using the geomean(recruits/spawner) for the first twenty-five years of each run. The model was initialized at the current (low) abundances, and under the alternatives, abundance tends to increase over time and so “low” abundance tends to occur early in the simulated trajectory. The authors indicated in their response to the ISAB that this followed the precedent of the Interior Columbia TRT.

What guarantees that those years are characterized by sufficiently low fish abundance? Would it be better to take recruits/spawner for the 25 cohorts with lowest spawner abundance? While the method of computing productivity is inconsistent with the TRT criteria, which define productivity as the "intrinsic productivity" of the population (p. 16 in McElhany et al. 2006), it follows an earlier precedent. As a result, they will almost always underestimate productivity relative to the TRT criteria. However, the authors only need an “index” of productivity from a single run of the LCM to create the logistic regression to estimate P(QET), and so it really does not matter what is computed. This could be clarified in the report so that a reader is aware that a consistent “bias” is acceptable in estimating the P(QET).

Figure 1.4 shows how to convert the (predicted) P(QET) to the corresponding VSP score. This is a simple linear interpolation. This piece-wise interpolation is inconsistent with the TRT approach, which defined integer-valued scores corresponding to ranges of persistence (1 - extinction) probability, i.e., a step function (cf. Fig. 9 in McElhany et al. 2006; ISAB report Figure 9, below).

The TRT made clear that P(QET) should be estimated for a naturally reproducing population, without contribution from hatchery fish. The TRT also describes a "viability curve" approach to relate abundance and productivity to P(QET). The modeling team should determine if the model analyses in this report are consistent with the TRT criteria. If not, the analyses should be revised or the report should explain why the analyses are valid.

First, comparing their curve (Fig. 1.4⁸) with the values in Table 1, the authors always overestimate the VSP score relative to the TRT's definition as shown in ISAB report Figure 8, below, of the TRT criteria (in blue) on top of Fig. 1.4 (note that the point at 0.5 extinction risks should be at 0.6).

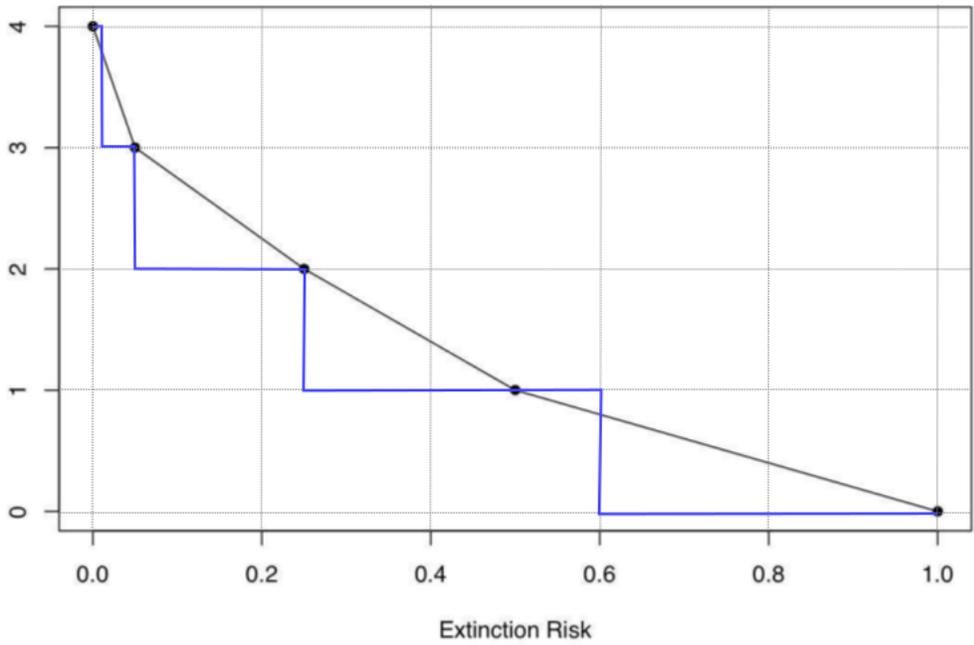


Figure 8. Comparison of piece-wise linear interpolation used in the NOAA report (gray curve) with the TRT step-wise interpolation (blue curve).

While use of the piece-wise linear interpolation will bias the absolute model scores relative to the TRT step-wise interpolation, it would have less effect on the relative rankings of alternatives. Additionally, this would imply in Figure 1.8 that this graph should only have four bars if they properly follow the TRT definition of VSP scores. The ISAB is not unduly concerned about fraction VSP scores, but the authors of the report should make it clear that these are inconsistent with the TRT scoring system and overestimates the score so that readers properly interpret the results when they examine the raw reported VSP scores.

Second, it is unclear if or how the QET analysis has been adjusted to remove the influence of hatchery fish. The definitions of P(QET), abundance, and productivity (p. 18) do not specify if abundances used are natural-origin or total. This may be particularly problematic for estimates of productivity and P(QET). Productivity is estimated as recruits (returning spawners) per

⁸ Figure and table numbers in this subsection refer to NOAA's report; ISAB Figures 8 and 9 are specifically identified as ISAB report figures.

spawner. Ideally, this would be computed as natural-origin recruits per total spawners, perhaps with an adjustment to total spawners for the reduced reproductive success of hatchery-origin fish. The legend of NOAA Figure 12.1.11 that an equation for recruits/spawner shows it is computed as $\text{NOR recruits} / (\text{NOR} + \text{Hatchery Spawners})$; additional details should be provided earlier. $P(\text{QET})$ is estimated directly from the simulations as the proportion of runs where abundance (is this total or just NOR) drops below the QET. Even if they use natural-origin abundance in this, the dynamics will be affected by past hatchery supplementation. To properly estimate $P(\text{QET})$, simulations excluding hatchery production should be used; it is not clear this was done. To the extent that hatchery production is mixed into the natural populations, not properly accounting for that mixture will underestimate $P(\text{QET})$ and overestimate abundance and productivity. This could be a problem for several Chinook populations with high hatchery influence. Moreover, the assumption that there is no straying within or among rivers may be important here. "In general, natural and hatchery origin adults return to their natal river and are allocated to the specific reach where they were born (straying to non-natal reaches or other rivers is not considered in the current model." P. 13. Does this mean that all hatchery-origin fish are assumed to return to their natal hatchery and not spawn in rivers? This should be clear; in some cases, there can be substantial exchange between a hatchery and the spawning grounds of that river.

Third, in step 1 of the short-cut to estimate variability of $P(\text{QET})$, a quasi-linear bivariate logistic regression of quasi-extinction (a binary indicator of whether a run dropped below QET) against estimated abundance and productivity for each model run is used to generate $P(\text{QET})$ response surface similar to a set of viability curves (Fig. 1.5). However, the model they use is inappropriate for modeling extinction risk and generates theoretically impossible $P(\text{QET})$ values. To illustrate the problem, compare their example (Fig. 1.5) with Figure 9 from the TRT (McElhany et al. 2006) below. Both purport to model extinction risk for Chinook from the Willamette. However, while the TRT curves increase rapidly as R/S declines toward 1, the curves in the NOAA model report do not; in fact, they continue to show low risk at R/S values well below 1.0, which would appear to be theoretically impossible for a natural population. Any population that fails to replace itself cannot persist indefinitely, and any with R/S below 0.7 are almost certain to approach zero within 100 years (ca. 25 generations for Chinook). Is this an issue with the definition of the recruits per spawner, which appear to be defined as $\text{NOR recruits} / (\text{NOR} + \text{Hatchery Spawners})$ in Figure 12.1.11? If so, values of 0.7 may just reflect a large number of hatchery spawners and the R/S values then are not easily interpreted. It is confusing that the model is generating runs (Figure 1.7) of abundances in the thousands with R/S well below 1, which then must imply that LCM itself is not reasonable for estimating extinction risk, perhaps because of the issues with hatchery influence mentioned above.

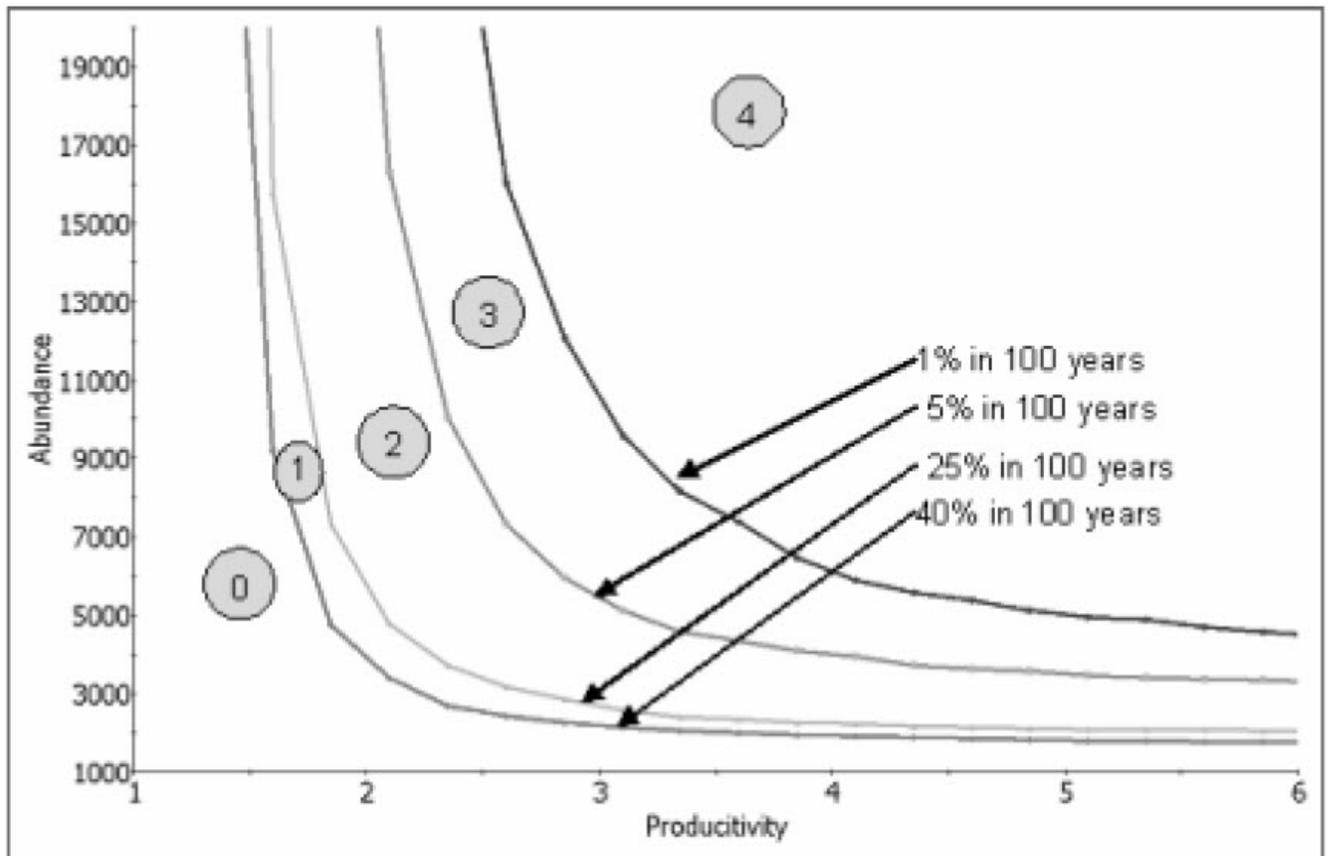


Figure 9. Viability curves showing relationship between risk levels and population persistence categories (example based on Chinook curve). Each curve indicates a different risk level. The numbers in circles are the persistence categories associated with each region of the chart (i.e., area between the curves). A population with a risk category 0 is described as a population that is nearly extinct and population with a risk category of 3 is described as “viable.” Taken from Figure 9 in McElhany et al. (2006).

Lastly, the first VSP score has such a high weight and is related to $P(QET)$, it is important that this is computed properly. The $P(QET)$ response surface is estimated from a single sub-run of each parameter set using an additive model (equation in the middle of page 20) based on a (power) of the recruit/spawner and abundance. The additive model places very strong constraints on the response surface (i.e., parallel lines on the logit-scale). With only one sub-run per parameter set and a very “tight” region where the model runs lie (Figure 1.7), there may be insufficient “sub-runs” to detect a non-additive structure. When Figure 1.7 is examined, there are actually no runs where the recruits/spawner is near 0.4, so the contours in that area are “imaginary.”

At the very least, a model that allows for an interaction between the recruits/spawner and abundance, or a response surface fit using thin-plate splines should be explored. These models will likely require more than one sub-run per parameter set, but hopefully not 1000 sub-runs, so computational burdens should not be extreme.

Figure 1.7 is very important to understanding the output, but no similar figure appears for the other alternatives. Given the high impact of recruits/spawners on population viability without large hatchery infusions, viable populations likely require fairly high numbers of recruits/spawner (e.g., greater than 0.8). Year-to-year variation in recruitment is large in salmon populations, so it is possible that boom years could alleviate generations of bust years, but presumably more extinctions should occur at low recruits/spawner. Figure 1.7 show that a median VSP around 2.6 with a predicted geomean recruits/spawner around 0.7 and (from Figure 1.4), a P(QET) in the 15-20% range. This would seem to imply that the population almost never replaces itself, regardless of abundance, under the baseline scenario. Does this then imply hatchery population is actually sustaining the population, which is ironic considering PHOS inherently lowers the VSP score.

The overall VSP score gives a weight of 4/6 to the P(QET) component and 1/6 to each of the Diversity VSP and Spatial structure VSP. The document does not explain the origin of these weights, but additional information was presented in their response to ISAB questions. According to the authors, this weighting scheme was developed by the TRT to reflect their relative importance for abundance and productivity. Details should be included with the report.

It would be helpful to give an example of the computations (either from an existing scenario or a made-up scenario) showing how the LCM trajectory for a single model run from a parameter set gives the various VSP scores.

Figure 1.8 appears to show variation in the VSP score for P(QET). Similar figures exist for the VSP components that are summarized in the median/percentile plots. This report is unclear on which sources of variation are captured. Each model run chooses a new set of parameter values, so this captures some of the uncertainty in some of parameter values. However, other parameters are assumed fixed (e.g., ocean survival), some are deterministic (e.g., FBW inputs), movement between life stages is stochastic (e.g., an 80% survival probability does not imply that exactly 80% of fish survive), and changes in parameter distributions from the calibration process all introduce additional variability into the LCM outputs. The report does not clearly explain which sources of stochasticity are included in the LCMs. Consequently, the apparent variability in VSP scores is likely understated to an unknown degree. The authors need to discuss this in more detail.

There appears to be too much reliance on the total VSP index, which is a summary of outputs. The figures in the Summary seem to show no clear alternative that performs best, and the perhaps alternatives 1 and 4 have promise, but the degree of responses in the components (e.g., abundance versus diversity) of the VSP index is not discussed. Later in the report, the components of VSP are reported, which is good. Graphs of actual model summaries may also be helpful, e.g., show the distribution of the P(QET) directly.

Minor Comments

p. 13, para 1. The document indicates annual time steps are using in the models, but dam passage uses "monthly estimates of fish of each life stage." How are the different time steps reconciled?

p. 13. Straying of adults is not currently modeled. Is information available to know if this assumption is reasonable? This implies that if a sub-population goes extinct, this available habitat will not be recolonized. Is this a reasonable assumption? (Refer to an earlier comment for a reference about straying).

p. 13 para.2. Are the three life-history strategies for each species sufficient? Steelhead, in particular, are known in other basins for high plasticity in juvenile rearing strategies.

p. 14, para. 2. The discussion of hatchery production seems to be for Chinook salmon only. This should be clearly stated. The final sentence explains that there is no significant winter steelhead production, which seems to assume no issues with summer steelhead production in these basins. Is there evidence that summer (hatchery) and winter (natural) steelhead do not interact?

p. 15. Hatchery releases were decreased when naturally produced adult abundances increased. The basis for this relationship is not clear. Is the maximum hatchery production level fixed?

p. 20. The equation used to estimate P(QET) for a single model run from a parameter set has several exponents that are not defined. Similarly, in the description of the logistic regression, the authors state that 0 was used if the run fell below QET and 1 was used if the run did NOT fall below QET. This would imply that the estimating equation estimates P(not QET). This may be an artefact of how *R* codes the two levels in a logistic regression.

p. 23. Is it appropriate to assume independence when calculating combined variance of the overall VSP score? It is not clear why this is even computed since you have a total VSP score for each model run and you can find the variability directly over the model run.

Review of Chapter 2. Juvenile Capacity

This is the first of several chapters that describe parts of the LCM in more detail compared to Chapter 1.

Table 2.1 and 2.2 show estimates of spawner capacity, which implies egg capacity. Juvenile capacity was thought to be much larger than could be currently produced and so is not modeled. Is this a reasonable assumption? This assumption seems reasonable for most of the models, which are based on adult abundance that has been calibrated to observed abundance. It probably does not matter where the capacity bottleneck occurs in the life cycle. However, local juvenile capacity could affect their diversity (based on juvenile life-history pathways) and spatial structure scores, but the report does not explain how those are estimated. The new estimates of habitat capacity (Table 2.1) are dramatically lower than previous values. This seems more than an adjustment as more data becomes available. What does this affect? Is the model sensitive to this?

The statement regarding Spawning/Incubation Habitat that “More contemporary surveys by R2 Resource Consultants are considerably lower...” should be re-examined. Presumably, it is the estimates from the surveys that are lower, not the surveys themselves, which might imply that they were conducted lower in the river system, lower in spatial detail, or some other survey attribute. More importantly, if “... we [the NOAA model, presumably] set capacities at 50% of the pre-dam estimates...”, how does this compare to what R2 surveys indicated? The only comparison implied is between the historical condition and the R2 surveys, not between R2 and NOAA’s 50% value. Or, if R2’s surveys indicated a 50% reduction, this should be clear. Given that “spawner abundance ... rarely produced sufficient numbers of juveniles to have parr rearing capacity limit population growth” it would seem that spawning ground capacity assumptions and adult returns are important. A similar situation was reported for steelhead, leading the authors to conclude that it was safe “not to include juvenile rearing capacity in the LCM for both Chinook salmon and steelhead.” The implications of this conclusion for adult abundance and spawning ground capacity should be made clear, as this seems important.

It would be helpful if Table 2.1 reported the egg to fry survival rates accompanying the “Mean Egg Capacity.” Also, presumably the “adult spawners” includes males as well as females. The estimated per capita fecundity from the table is about 2250, which is about half what one

would expect from Chinook. This should be made clear in the caption. Is it correct that “egg capacity” is twice the number of adult spawners x 2250 in this case? Was the capacity for eggs set first, and then the number of adults back-calculated from it?

Table 2.2 uses the same egg production per capita for steelhead as was used for Chinook. This is not what the UBC team used, though the difference is not great. It would have been good if all teams had used the same values.

Review of Chapter 3. FBW

The alternatives that are considered involve alterations to dam operations. The FBW is applied to each water year for each alternative, and a single DPE and DPS are obtained for each water year for each alternative (Table 3.1; Figure 3.1). The different alternatives also affect downstream conditions (TDG and temperature, which are modeled; habitat and floodplain inundation, which are not modeled).

It is unclear how fish are passed to different outlets if empirical data do not exist because the infrastructure (e.g., floating collection structures) does not exist? What is the source of the distributions in Table 3.1? Since the outcomes are so sensitive to these parameters, but also vary even within a reservoir and species, this seems like an important source of uncertainty. Page 34 says fish are apportioned to different outlets based on proportions of flow, but there is considerable literature that suggests that fish are not just passive particles. How widespread is this assumption in the LCMs? The authors provided additional information in their response to the ISAB:

“Fish movement through various outlets (Table 3.1) was determined by the Fish Benefit Workbook as developed by the Corps. In part, our understanding is that route allocation is determined by depth to the outlet, absolute and relative flow through each route, characteristics of specific structures, and the species and life history stage of the fish being modeled. These numbers and distributions were supplied by the Corps, and the LCMs used these FBW outputs (DPEs and survivals; e.g., Fig. 3.1, p. 29 and see the appendix) to specify the fish experience interacting with and moving through the dams.”

It appears that the FBW is treated as a “black box” that provides the necessary inputs to the LCM models. Refer to our review on the FBW above in this report.

The Alt3b alternative have a very flat DPS compared to the other alternatives. Why? Is this realistic? Is this a flaw in the FBW? Figure 3.1 appears to have no stochasticity even though

water conditions vary considerably from year-to-year. Because the FBW outputs appears to be the driving factor that differentiates the results of the alternatives, does this imply that the observed variability in the VSP scores is understated?

The discussion of the delay or forced suboptimal passage in the FBW seems to be an important limitation in representing the drawdown/operational alternatives. Could the effect be large enough to explain the low performance of the operational strategies? There are a number of issues. For example, the FBW estimates average daily reservoir conditions for monthly time steps, but flows can change substantially within a month because of operational schedules (e.g., drawdown). Also, it does not seem realistic that fish will wait several months before trying to migrate again. Is there any empirical information regarding how many fish will quickly repeat a passage attempt versus waiting until the fall to try again? This potential bias seems like an important point of discussion/for improvement. The authors response to questions from the ISAB indicates that the FBW is treated as a “black box” that provides the necessary inputs to the LCMs:

“We believe that the modeled success of operations may be underestimated for the reasons stated. It is our hope that current operational passage studies under the court injunction may provide some ground truthing for future FBW estimates.”

The importance of the FBW outputs in driving the differences in LCM predictions of response to alternatives should be investigated. One easy approach is for the authors to provide a summary of rankings solely using FBW and from the LCMs to see if they are congruent. If not, explain why not.

Review of Chapter 4. Juvenile post-dam passage

A logistic model is used to estimate TDG mortality for alevins and juveniles passing the dam. The two equations are applied to FBW output to get a single number for each alternative for each water condition. Are these parameters treated as fixed in the LCM runs? Is demographic stochasticity applied for any survival computation during an LCM run?

Figure 4.2 shows that TDG mortalities are negligible for the first day, and only increase dramatically after 2 days of exposure. Given the movements of the fish downstream, how long are they exposed to lethal and sublethal levels? Put another way, how do we go from Maule’s lab study to the field? What assumptions of movement were used to produce the lower mortality estimates?

Review of Chapter 5. Ocean survival

In this chapter, the authors use climate indicator regressions to reconstruct past survival histories. The authors develop a model for the survival in the initial year in the ocean (s_3) and then assume that survival in subsequent years is a fixed 0.8 for both Chinook and steelhead.

An estimate of s_3 was based on SARs and proportion of each age class returning. They adjusted the s_3 estimates to account for hatchery vs wild survival. The authors assumed that wild fish have about twice the marine survival of hatchery fish. Some estimates give wild fish a bigger advantage under unfavorable conditions (e.g., Beamish et al. 2012 in British Columbia) whereas other studies report little difference (Woodson et al. 2013 in California). The 50% value is relatively close to those given in the classic study by Raymond (1988), which reports wild vs hatchery Chinook as averaging 2.3 vs. 1.3%, and steelhead as 3.0 vs. 1.6%. The authors should cite these and similar papers as the basis for their value, and also acknowledge that using a fixed value is a strong assumption for a variable that shows high variability.

This estimated s_3 was used in a (empirical) logistic regression against the PDO, upwelling index, river flows. The authors used all possible regressions methods and ranked models using AIC, and such. They then predicted s_3 over all ocean cycles. Stochasticity was included in s_3 using prediction intervals for s_3 from this top supported model. As well, each model run started at a randomly selected set of ocean conditions and set of water years for each simulation. This assumes that ocean conditions are independent of water years. Is this a sensible assumption (see page 52 in the NOAA report for more details).

There are some potential issues with this approach. First and most important is that these relationships are known to be ephemeral (e.g., Litzow et al. 2018; Wainwright 2021). Models fitted to short time series should not be used for longer-term reconstructions. Second, they choose best-fit models but do not indicate if these models are significantly better than alternative models, which might result in different survival estimates in out-of-data years. A table with statistics for the various models would help answer this. A multi-model approach (e.g., Rupp et al. 2011; Burke et al. 2013) might give more meaningful reconstructions, as might use of a simple time-series (ARMA) model without tying survival to particular climate indicators. A third potential issue is that the approach assumes that different populations within a species respond to climate variables independently, even though they all experience the same ocean at the same time. This results in inconsistent model selections among populations. As an example, models selected for different Chinook populations variously use the PDO in April, May, and June, despite the PDO being a decadal-scale indicator with meaningless month-to-month

variation. This in and of itself suggests that the model selection process could lead to spurious results.

The report is unclear how stochasticity is included. For example, did they use the mean survival for each year or draw from the total prediction distribution for each year? These approaches would result in large differences in uncertainty of results. If they took the latter approach, distributions are broad enough that the regression issues may be unimportant.

Minor point

p. 52. The authors may wish to update citations about synchronicity to include more recent studies, e.g., Chasco et al. (2021).

Review of Chapter 6 harvest

Ford (2022) reports total exploitation ranging from about 20-40% over the relevant period (Fig. 54). The value of 10.2% was not found in Ford's report but was presumably calculated in some manner using only ocean fisheries. The NOAA Willamette report is hard to follow, as it indicates 10.2% taken in ocean fisheries, 10% incidental take in catch and release mainstem fisheries, 12.2% incidental take in catch and release Willamette River fisheries, and incidental mortality of 14.7% and 40.0% in Columbia River tangle net and large-mess gillnet fisheries. How do these values add up? Some fisheries will take place after others (e.g., ocean fisheries will occur first), and presumably these are values based on the fish contacted by each fishery? Is the 21.1% average for freshwater impacts (just above Table 6.1) added to the 10.2% from the ocean for a total of 31.4% on average?

Minor comments

p. 53. The modelers set the Chinook ocean exploitation rate to 10.2% overall exploitation probability. How was stochasticity included in this?

p. 54. Chinook freshwater exploitation rates. How were they combined and has stochasticity been included in this?

p. 55. How was stochasticity included in steelhead exploitation rates in ocean and freshwater?

Review of Chapter 7 Freshwater adult mortality

The authors develop a logistic model for en-route mortality as function of environmental covariates (discharge, water temperature, air temperature, precipitation). All possible logistic

regressions were fit with model ranking using AIC and such. The best model had water temperature and discharge as shown in Figure 7.1. How is stochasticity handled here?

Minor comment

p. 57. The modelers needed to estimate water temperature in earlier years. How was this “error” in variables handled? More details are needed here.

Is it possible that some of the apparent en route mortality results from fishing that is not sampled or accounted for? Further discussion is needed.

Review of Chapter 8. Pre-spawning mortality.

The authors created an empirical logistic model to estimate pre-spawning mortality (PSM) as a function of pHOS and water temperature. Bowerman et al. (2018) paper reported higher PSM with higher pHOS, but this result may be related to differences between the arrival times on the spawning grounds for hatchery-origin fish as compared to natural-origin fish. For example, hatchery-origin salmon often return to spawn earlier than wild fish and may thus experience warmer water and higher PSM rates.

The authors recognize that water temperatures are only directly measured in recent years at USGS gauges. They develop a model to predict gauge temperatures as a function of environmental variables. The models performed well except at Foster Dam and Fall Creek Dam where dam operations have larger influence. It is unclear why they did their own temperature modeling instead of using the CE-QUAL-W2 results provided by the Corps. Perhaps that dataset does not cover the spawning areas?

The authors used an annual temperature that was calculated for July 01 to Sept 15 to represent “summertime” water temperatures (p. 53). We understand why they did this but also wonder what the effect may be on the model results because temperatures and flows change considerably between July and September in some basins. They demonstrate in Figure 8.3 that there appears to be a bias towards colder water in their model for that same time period, for which they corrected their model with a “delta” approach. Should that delta vary across basins or water years?

This model used the “delta” methods to impose temperature changes in other years. Did this model use different temperatures than the other models the ISAB is reviewing? This approach is another example of where decisions and assumptions have to be made with a model, but it is not clear how important it is in determining the results.

Minor comments

p. 66. A table of estimates is provided, but there is no visual assessment of fit, which would be helpful.

Review of Chapter 9. Hatchery processes

The authors developed a model to predict fitness of naturally spawning hatchery fish as function of pNI. It is not clear if this is based on “data” or “expert opinion.” Again, how is stochasticity included in the LCMs?

Table 9.1 revealed that only one population (Clackamas River above dam) is numerically dominated, on the spawning grounds, by natural-origin spring Chinook (97%).⁹ One other has 57% natural origin fish, and the others are around or below 25%. Is this consistent with the assumption that there is no straying within or among rivers? This seems to indicate considerable exchange between natural and hatchery origin fish in these rivers.

Review of Chapter 10 Calibration

While each sub-component of the LCMs may be grounded with empirical data and produce realistic results, there is no guarantee that the full models will produce sensible results. The authors used a method (Approximate Bayesian Computation, ABC) to “calibrate” the LCMs to data collected on the system.

Unfortunately, the data available for calibration are rather sparse. There are only 12 years of data (2006-2017) for three Chinook populations. There is very little data for the other Chinook populations or for steelhead. In the response to the ISAB, the authors gave additional information on why so few years of data are available for calibration:

“We had very few data to represent the ‘current’ or ‘baseline’ period (i.e., spawner abundance over a handful of years; this is related to Question # 8 below). Furthermore, dam operations have changed in the recent period of record of spawner estimates used for the calibration, which may or may not be reflected in the NAA, the alternative that was used for the calibration. We note that we know of no population-level estimates of

⁹ Note that the citation of Ford et al. 2012 in Table 9.1 appears to be typo and should be Ford et al. 2022. Biological Viability Assessment Update for Pacific Salmon and Steelhead Listed Under the Endangered Species Act: Pacific Northwest.

juveniles from any of the basins. Having an additional abundance estimate from a juvenile life stage would add a substantial amount to the calibration process. “

And

“We did not use all of the available years of data because of major changes in dam structure and/or operations. We went back in the time series as far as we thought reasonable without introducing substantial variation due to changes in population abundance surveys, dam operations, improvements in passage, etc. For example, in the North Santiam River there were changes at Bennett Dam that altered adult migration and operational temperature management at Detroit Dam that likely affect incubation success and adult prespawning mortality. Further, systematic spawner surveys were conducted in the four primary tributaries for about 10 years above and below the dams up until 2019. Additionally, there are no accurate estimates of hatchery contribution to natural spawners prior to 2002.”

This information should be included with the report.

Only a few parameters were chosen for calibration (those parameters not informed by exterior studies). According to the authors' responses to the ISAB, between six and eight parameters were included in the calibration process, depending on the LCMs. The choice of parameters to include in the calibration arose from uncertainty about parameters that were discussed in the expert panel process. The decision to discard was related to how much information was obtained from the calibration process: if the approximated posterior distribution resembled the prior distribution, then it was excluded, and the calibration process was repeated. There were many parameters that could have potentially been included; additional data for calibration of the whole model version is needed.

In the ABC process, a large number (50,000) of parameter sets are (randomly) generated. For each parameter set, the LCM is run and a comparison of the distribution of simulated spawner numbers vs the distribution of known spawner numbers is made, and a goodness-of-fit statistic is computed. Finally, the best fitting parameter sets (best 1% as measured by the goodness-of-fit statistic) are retained and this provides a posterior-like distribution for the parameter values. For populations with insufficient data, trial and error were used to adjust the LCM output until they appeared sensible, but no details are given on how this was done.

The KS test to compare the simulated distribution with the real distribution ignores all dynamic structure in the predictions, so it is only a test of whether the 100 years of predictions have

roughly the same mean and range as the short period of observations. This is a very weak form of comparison of a complex dynamic model. For example, it does not provide any information how well the model will reproduce the other VSP outputs: productivity, pHOS, life history diversity, spatial structure. The authors should report the full distribution information (box plots, percentiles) from the calibration method.

The 100-year vs 12-year comparison may be problematic if the 12 years were at the high or low end of conditions affecting salmon. This would tend to bias the calibration high or low.

Alternatively, it is not clear why they did not compare the predictions for the same set of years as the data. Or the calibration process could have focused on fewer parameters from the start, such as those parameters with little prior information and high sensitivities.

Some additional information is needed. For example, the authors dropped parameters from the calibration process because the data appeared to provide no information on their value, but a list of the dropped parameters is not provided. A table showing all of the parameters' values prior to calibration and after calibration would be helpful to understand which parameters are most affected by the calibration process. For example, Figure 10.2 could be modified to show the prior and calibrated distribution. In addition, it is not clear from Figure 10.2 if there were only four parameters that were calibrated for this population model.

For some parameters that were dropped from the calibration, their values were fixed rather than drawn from a distribution (page 80). Why did they shift to fixed values here, rather than just using the prior? This would seem to reduce variation in the final VSP scores, depending on which parameters these were and what their sensitivities are.

Note that accurate calibration is not necessarily critical for relative comparisons of the EIS alternatives, unless calibration bias causes the results across alternatives to stack at the low or high limits of the output scores, which would make the analysis insensitive. This is most likely to be a problem for the abundance and productivity scores, as extinction risk and its translation to the VSP score are nonlinear.

Review of Chapter 11 Life Cycle Population Descriptions

In this chapter, the authors describe in general terms the output from the respective LCMs.

The authors briefly discuss how the NAA simulation may not match present-day conditions that generated the data, but they used it anyway because they always compared alternatives to

NAA. This needs to be further deconstructed, evaluated, and discussed. This could have major implications on how the results are interpreted.

Review of Chapter 12. Life Cycle Modelling Results

This is a final comparison of the VSP scores under the various alternatives. As noted previously, the graphs and descriptions could be improved. For each sub-population, the various alternatives are compared. It would be helpful to also compare across populations for any of the output criteria. In this way, “consistency” of the alternatives across the various populations could be assessed and a critical evaluation of major differences performed. Additional comparisons of the underlying metrics (e.g., P(QET)) are also presented.

There are also some apparent inconsistencies that need explanation. For example, Figure 12.2.4 shows that P(QET) is 0 for all alternatives. Yet Figure 12.2.5 does not have the VSP score for abundance and productivity at 4 for all of the alternatives? This is especially true for the NAA.

Some unusual dynamics are predicted and may be correct, but these need to be further explained. For example, consider Figure 12.1.11. The model is in steady state, but alternatives appear to start at different initial conditions. Will this not affect the estimates of productivity based on the first 25 years of model output? Additional discussion is needed about these dynamics.

Because the alternatives are directly compared using the VSP scores and not directly using measures such as mean abundance, an alternative that performs only slightly better in the VSP components could be rated highest, but in actuality it is essentially the same as the other alternatives. Consequently, alternatives should also be compared in native units (such as mean abundance) for the alternatives so not only is the best one identified but one can easily determine how much better it is. The authors present some of these graphs (e.g., Figure 12.1.2 for total NOR spawner abundances and Figure 12.1.4 for the individual reaches; Figure 12.1.3 for P(QET); Figure 12.1.7 for pHOS), which is laudable, but more discussion is needed on comparing and contrasting the results from the VSP and the underlying units).

Minor comments

The authors changed the type of graph (e.g., bar chart, rounded bar chart, tick marks at end of whiskers) without any explanation. As far as we can tell, the same percentile plots could be used for most of the graphs. This would make it easier for the reader to interpret the graphs.

Review of Chapter 13. Conclusions

The authors conclude that the differing FBW outputs among the alternatives was the major source differentiating among alternatives. The two key metrics of the FBW are the DPE and DPS. Do you get similar results to the VSP score by a simple comparison of these two metrics across the alternatives? If so, then what value added do the LCMs provide and how does this new information help in the decision making?

We are also concerned about drawing too strong of conclusions regarding some of the operational strategies (e.g., drawdown) given the limitations of the FBW noted on pages 34-36 and 136. This is particularly relevant given that the population estimates so closely reflect the FBW passage efficiencies. They are clear about this on page 136 as well, but they do not go far enough to say that the FBW needs to be updated to address the timing issues in the model. They need to clarify if they think that the issue does not affect the model result. The data we have seen for the Fall Creek drawdown shows impressively high efficiencies. Regardless, this report does not demonstrate whether this a model issue or an operational issue. Again, the authors indicated in the response to questions from the ISAB: "It is our understanding that the USACE is looking into ways to improve the FBW - ... we hope empirical studies will provide some insight into the extent of this problem." The more general issue is that the FBW is treated as a "black-box."

The discussion of limitations of the study was very informative.

The authors discuss population sustainability, viability, and recovery abundances in absolute terms. The modeling process, especially given the paucity of calibration data and potential calibration biases, does not meet the standards for Population Viability Assessment models (e.g., Ralls et al. 2002), and thus cannot support these conclusions. Rather, conclusions would be better expressed as relative comparisons across alternatives, telling us what alternatives produced, for example, higher or lower sustainability relative to NAA.

Minor comments

p. 135, para. 2. "Comparisons across the NAA and six alternatives for all ... populations ..."
Again, a summary graphic would be very helpful with this discussion, replacing the footnote with something more readily understandable.

p. 136. Is it reasonable to assume that transport has "minimal effect?" Which studies is this assumption based on (e.g., from the CSS reports)? And how is this "minimal effect" implemented in the models, or was it deemed to be so small and simply ignored?

p. 138. The authors comment again that the fish data applied for the NAA were not collected during operations that reflect the NAA. This seems important, though it is not clear how much so. Is it possible to rerun the models with ResSim operations that reflect the operations when those data were collected?

Review of Appendices

There are some very large differences in estimates produced by the FBW (e.g., Figure 14.1.3). Are such differences realistic? What are the causes of these large changes – e.g., abnormally high or low spill?

Editorial Corrections

p. 5. para 1. Spell out NAA on first usage.

p. 5. para 2. "... highlight the ability of management efforts." Incomplete sentence.

p. 12. Figure 1.1. Legend has three labels that are coded rather than descriptive -- "willamette_river" is self-explanatory, but "newhuclcw27" is not.

p. 13. para. 2. Is there a difference between the life-history terms "sub-yearling" for Chinook and "parr" for steelhead? Both terms seem to refer to downstream migrants in the hatch year.

p. 15. How does the NAA not capture current dam configurations and operations?

p. 18. last paragraph. References to McElhany et al. (2000) should be to McElhany et al. (2006) as shown in the reports reference list, unless the earlier VSP report (below) was the intended citation, in which case it should be included.

p. 12. Figure 1.1. The last two legend entries should either be removed or explained. Can the watersheds with the Chinook and Steelhead populations be identified in the legend or on the map with a symbol?

p. 14. Figure 1.2. Define LH1, LH2, LH3 in the legend to make the figure self-contained.

p. 22. Figure 1.7. This is presumably for the McKenzie Spring Chinook population coming from Figure 1.6? All figures legends should be reviewed to ensure that the population it refers to is

given. Mask Figure 1.7 (and similar figures) to show only the contour lines around a convex hull (or slightly larger than the convex hull) of the points from the runs for an alternative so that the “interpolation” space is well defined.

p. 22. Figure 1.7 (and elsewhere). It will be difficult for color-blind readers to see the red dot, so use a different (larger) symbol.

p. 22. Figure 1.7 (and Figure 1.6). It will be helpful to note, in the captions of figures 1.6 and 1.7, that a VSP score of 1.0 corresponds to a 50% extinction risk, and higher scores indicate lower extinction risk. Otherwise, the meaning of the isoclines is not self-evident.

p. 29. Figure 3.1. The top graph is labeled survival, but perhaps label it DPS to match the text.

p. 38 and 39. Recode the model to avoid intercepts terms of <-20 since these are numerically unstable.

p. 54. In Table 6.1, what exactly is “percent impact”? Percent of surviving fish estimated to be killed indirectly and directly? This should be stated explicitly.

p. 56. Second paragraph under 7.1 missing reference to a Figure number.

p. 83. Figure 10.2. and following figures. Better labelling of the parameter is needed. For example, “Fry.avail.pass.CGR” is not “fry available to pass CGR” since it appears to be a proportion (values between 0 and 1). A similar comment applies for the other parameters.

p. 98. Consistency: here (and only here) “PrQET” is defined and used; previously “P(QET)” was used.

p. 99. Figures 12.1.1 and Figures 12.1.2 (and others). Why is the form the graphs slightly different here, e.g., larger squares for the median rather than a tick mark?

p. 100. Figures 12.1.3 (and others). Why has the format of the plot changed here to a bar chart?

p. 101. Figure 12.1.4 (and others). Again, why has the format of the figures changed, e.g. tick marks at the limits of the whiskers.

p. 103. Figure 12.1.9 (and others). Why does the format change to a bar chart here? The same percentile graphs can be used.

p. 104. Figure 12.1.11 (and others). The Y-axis is labeled the proportion of natural origin spawners to total number of spawners and is not the recruits per spawner as stated in the figure legend. This cannot be this ratio since values are outside the 0 to 1 range. If these are recruits/spawner, then how are these mapped to VSP scores without also considering the abundance since the P(QET) depends on both recruits/spawner and abundance? Why is the initial recruits/spawner so different in the first 25 years for some alternatives for some populations? X-axis not labeled (years?).

p. 107. Figure 12.2.5. Given that P(QET) rarely goes above 20%, perhaps change the Y-axis limits to avoid pushing the results in to the bottom axis.

p. 122. Figure 12.5.2. "955" should read "95%."

p. 125. Figure 12.5.9. The tick mark labels for the Y axis need to be in numeric rather than scientific notation. Check all the other figures as well.

p. 129. Figure 12.6.2. Graph has an extra "title" at the top that is not needed with a legend.

p. 130. Figure 12.6.4 (and others). Why does the format change to a rounded bar chart here? The same percentile graphs can be used.

p. 133. Figure 12.6.10 (and others). Why do the percentile ranges appear to be one-sided for some alternatives?

D. Review of Ecosystem Diagnosis and Treatment Models

1. Summary Comments

Strengths

- The EDT model has been used widely in the Columbia River basin, and several independent analyses have evaluated its projections.
- The model's spatial domain is relatively similar to the spatial extents of the NOAA and UBC models.

Limitations

- The EDT models are not full life cycle models. They represent the response of a single cohort to the available habitat. Abundance of returning adults does not affect the populations in subsequent years.
- Details of the models and numerous rules and indices are described only briefly, but details are available in other publications.
- The EDT models for the WVS PEIS are run for three pre-selected years (2011, 2015, 2016). The robustness of results would be increased with a broader selection of years, and if some of these years reflect likely future conditions in the WVS.
- No sensitivity analysis of these models is reported. Past studies of other EDT models have reported sensitivity that can inform the WVS application; however, model sensitivities can be influenced by the specific features of the basin and how the models are implemented.
- The spatial domain includes the mainstem Willamette River, but differences in flow were not represented as differences in depth, width, or floodplain inundation. Because of this limitation, the effects of the dams and flow management on downstream habitats and water quality in the mainstem Willamette River are not represented.

Recommendations

- Not all the specific rules used for the models of Chinook and steelhead in the Willamette are reported. These should be reported in an appendix or supplemental material.
- For future review and use of these results, the data and model version used for the analyses of the EIS alternatives should be documented and archived.

2. Background: EDT

The Ecosystem Diagnosis and Treatment (EDT) model (Lestelle et al. 2004) was developed through the collective efforts of many biologists, modelers, and land managers (both

governmental and non-governmental) to assess effects of management actions on habitat conditions, fish abundance and distribution, including the impact of hatchery management strategies on ESA-listed fish and harvest objectives. The Northwest Power and Conservation Council (NPCC) sponsored a project to develop a Multi-Species Framework and asked the EDT modelers to develop a version of EDT that could be applied consistently across all subbasins in the Columbia River basin. EDT was used to assess existing and future restored habitats in most of the NPCC subbasin reports for the Columbia River Basin Fish and Wildlife Program ([ISAB/ISRP 2004-13](#)). The EDT model was previously applied in the Willamette River and has since been updated with more recent information (ICF 2022).

EDT is a rule-based model to evaluate habitat influences on biological performance. Lestelle et al. (1994) noted that EDT is not a predictive model and was designed to “help assess uncertainty and the relative risks and benefits of alternative supplementation strategies.” As its name implies, the EDT model was originally designed to diagnose salmonid habitat limitations in a river system and compare alternative habitat treatments in terms of capacity and productivity (as defined for a Beverton-Holt production function). The models for Chinook salmon and steelhead predict population level estimates of capacity, productivity, diversity of productive habitats, and equilibrium abundance by scenario based on freshwater habitat conditions. Beyond traditional aspects of habitat, EDT also incorporates survival factors such as predation, harvest, obstructions, and entrainment/injury by water withdrawals, primary as fixed effects. The models start with "benchmark" parameters for each life stage of each species, representing capacity and survival under ideal conditions. These benchmarks are then reduced for each life stage and stream reach using rule-based interpretations of actual conditions (or projected conditions under a set of management alternatives). The models run for each year are independent from other runs for other years and based on the benchmark parameters. Returns of spawners do not affect the production of eggs for the subsequent generations.

The EDT model has been peer reviewed and implemented in many subbasins in the Columbia River system. Rawding (2004) found that spawner-recruit curves generated by EDT models were consistent with observed spawner-recruit curves for approximately two-thirds of the eight populations of steelhead, three of Chinook, and one of chum salmon in multiple tributaries of the Columbia River basin. The ISAB and ISRP reviews of subbasins plans recommended that EDT results be considered hypotheses to be tested because of the substantial uncertainty in the model estimates ([ISAB/ISRP 2004-13](#)). Steel et al. (2009) conducted sensitivity analyses of the EDT model and concluded that uncertainties in the input parameters create wide confidence intervals around estimates of abundance and productivity and also suggested that the model’s relative evaluation of the effects of habitat quality may be robust. Steel et al. (2009) reported that Bureau of Reclamation (Yoder et al 2006) found that EDT was relatively insensitive to flow attributes in the Yakima basin, which were important to their decision-making, so they

developed additional models to predict values for several habitat and temperature attributes incorporated in EDT that are most influenced by changes in stream flow. The EDT model application in the Willamette did not project influences of flow on habitat, but rather used estimates of numbers of fish passage at the dams based on the FBW. Similarly, McElhany et al. (2010) found that the EDT model might be more useful as a relative measure of fish performance than as an absolute measure because ground-truthing the projections of capacity and productivity with site-specific field data may not be possible in some applications. Roni et al. (2018) reviewed several approaches, including EDT, and recommended that its outputs should be seen as hypotheses of how a population functions and needs to be tested. An analysis of the application of EDT for the NPCC's subbasin reports in 2004 noted that a limitation of the model was the reliance on expert opinion to develop rules about key functional relationships with habitat (Hill et al. 2019).

The ISAB recently examined the use of the EDT model in the *Review of Spring Chinook in the Upper Columbia River* ([ISAB 2018-1](#)). The review found that the relative ranking of potential equilibrium abundance of adult spring Chinook salmon and steelhead estimated by the EDT models differed substantially from the relative ranking of available habitat (i.e., length and area) by the intrinsic potential models. The review recommended that a scientific evaluation of the species habitat rules would be prudent when EDT is applied in the rest of the Columbia River Basin. The ISAB's *Review of the Upper Columbia United Tribes' Fish Passage and Reintroduction Phase 1 Report: Investigations Upstream of Chief Joseph and Grand Coulee Dams* ([ISAB 2019-3](#)) also reviewed the model. The ISAB noted that the model was widely used in the basin and the representation of habitat conditions and integration of multiple sources of habitat data could provide useful representation of potential fish responses.

The EDT models require estimates of the effects of other factors throughout the full life history of the fish in downstream reaches and in the ocean, which are generally represented in the survival rate estimates. These survival estimates do not allow assessment of specific relationships in the lower Columbia River or Pacific Ocean (e.g., effects of avian, fish, or pinniped predators, total dissolved gas, commercial and sport harvest), nor address their variability.

Uncertainties are created in the EDT analyses where multiple sources of habitat conditions are obtained from different agencies and contractors and where missing habitat parameters or gaps in spatial coverage are filled (e.g., 3rd party models, aerial imagery interpretation, interpolation from comparable watersheds).

Laura McMullen (ICF International, Inc.) briefed the ISAB on the EDT model application for the PEIS of in the Willamette River in December 2021 and provided information and answers to the

ISAB's questions during our review. The ISAB reviewed a report submitted to the Corps (ICF 2022) and descriptions of the EDT models for the Willamette in Chapter 3.8 and Appendix E of the draft EIS (USACE 2022).

3. ISAB Answers to Review Questions

1. Model structure and development

- a. *Is the model structure appropriate for evaluating fish responses for the EIS alternatives?*
 - i. *Are the relevant processes (mortality, growth, reproduction, movement) included and are they represented in a way that encompasses the changes within and among the alternatives?*

The EDT models for Chinook salmon and steelhead are appropriate for evaluating fish responses for the EIS alternatives. The EDT models have several important characteristics that differ from the other population models (UBC and NOAA) used by the Corps for the Willamette EIS. The EDT models largely provide a habitat-based perspective of potential salmon populations for a single generation of salmon or steelhead for a representative year. They are rule-based models that use general relationships between habitat and fish abundance and survival to estimate the abundance of returning adult salmon, which have no influence on subsequent generations. The EDT models represent habitat relationships to a greater extent than the NOAA and UBC models, but they do not provide measures of population dynamics that result from multiple generations of salmon and responses to factors that influence survival. The single-generation context of the models makes them useful for examining short-term responses to flow conditions. Thus, use of EDT alone for the WVS PEIS would only provide part of the answers needed; the Corps used EDT along with the NOAA and UBC life cycle models in a multi-model analysis. EDT does not provide estimates of SARs or P(QET), which are important outcomes estimated by the NOAA Fisheries and UBC models. The measure of diversity from the EDT models is not the typical measure of life species diversity or history diversity used in ecology and fisheries science, which is important to recognize in interpreting its predictions of diversity.

The rule-based approach is developed using relatively general relationships and indices, and it does not include detailed representation of several major factors that can affect survival (e.g., predation, variation in ocean conditions, harvest). The collection of rules account for downstream and marine effects based on overall estimates of survival in those portions of the life history. Analyses of previous versions of the model found that it was insensitive to flow relationships (Steel et al. 2009), and the current version for the Willamette Draft EIS does not include differences in channel morphology and floodplain inundation that would result from different flow regimes. For the management alternatives evaluated in the Draft EIS, the EDT

models projected higher Chinook and steelhead abundances than the NOAA Fisheries and UBC models (Figure 10), and the relative differences in abundances between alternatives were smaller than for the other two models.

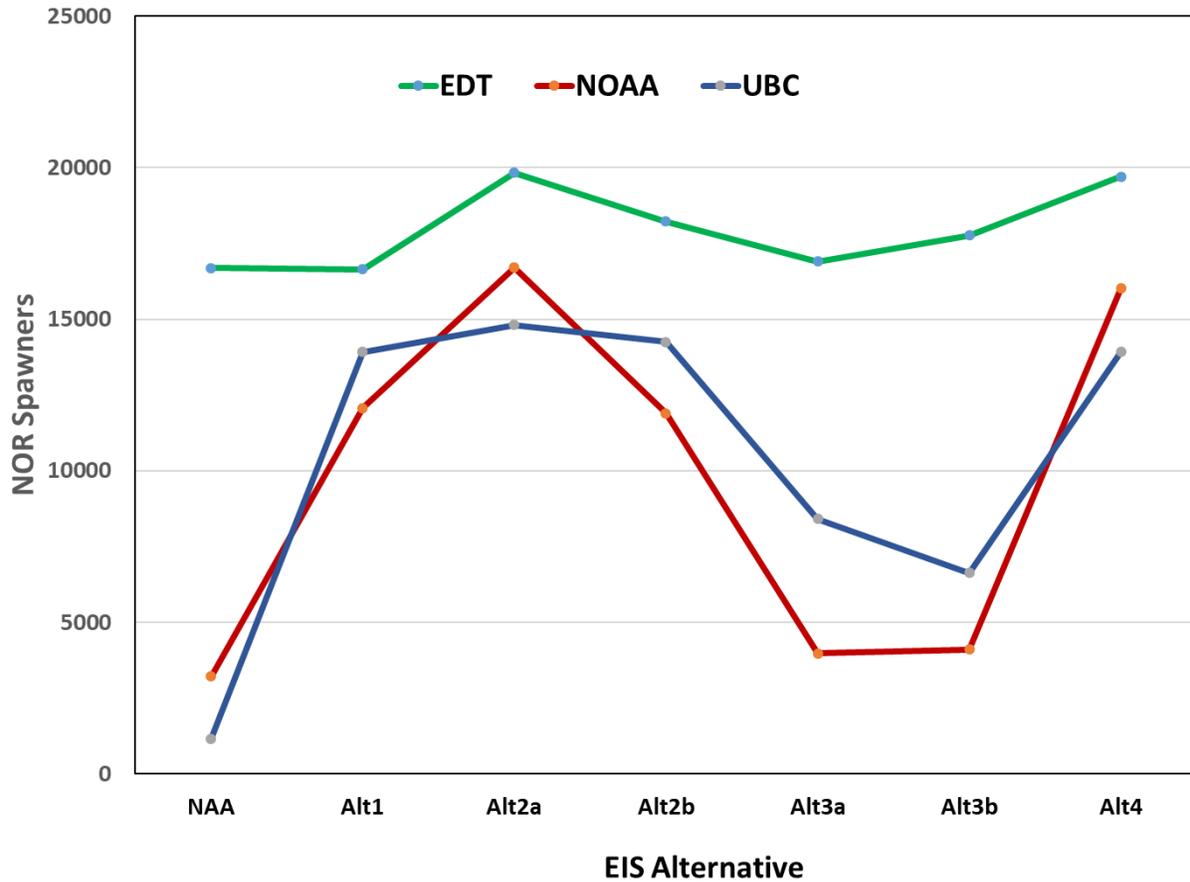


Figure 10. Comparison of the estimates of NOR spawners from the EDT, NOAA, and UBC models for the EIS alternatives (Appendix E in draft EIS, USACE 2022).

Detailed comments about model structure and development

Each trajectory of fish simulated in EDT starts with eggs and moves through life history to ocean and then return as spawners. The EDT report (ICF 2022) does not explain how the benchmark parameters are determined to set the initial conditions for each annual run. The EDT modeling team explained that the EDT model starts each run or trajectory with "benchmark" parameters for each life stage of each species, representing capacity and survival under ideal conditions. The universal benchmarks are values for the productivity and capacity of each species based on

literature values. These benchmarks represent optimal productivity and capacity under ideal conditions for different life stages. The values for the benchmarks for Chinook and steelhead in the Willamette EDT models are the same values reported in Appendix B of Lestelle et al. (2004). The values are general estimates from the Columbia River Basin, and their relevance to the local site-specific conditions of the WVS are unknown. The benchmarks have not been updated since 2004. While the models could be improved by updating the benchmark values, it would not likely greatly affect the application of the EDT models to the WVS PEIS, especially if predictions are compared as relative changes.

The EDT models run 3000 trajectories for Chinook and 2000 for steelhead. The EDT modeling team reported to the ISAB that the different numbers reflect differences in kilometers of spawning habitat for the two species, leading to similar numbers of trajectories per day per kilometer for the two species. The rationale for this difference in numbers of trajectories is not stated, and the basic idea of sampling at the same density for both species should be explained.

The report states that a single trajectory starts in one spawning location with a certain number of days in each stage. The EDT modeling team explained to the ISAB that the trajectories are chosen based on life-history strategies and represent the distribution of trajectories across life history strategies. Ensuring the collection of trajectories is a robust representation of how fish could use the habitats is critical to generating model results that are interpretable and can be used to compare among the management alternatives. The report could be improved by additional explanation of how trajectories are allocated spatially, temporally, and among life histories.

The convergence properties of the model predictions should be assessed to confirm that 3000 trajectories for Chinook and 2000 trajectories for steelhead are sufficient to generate output predictions. Table 2-2 on p. 2-6 shows distribution of trajectories in each life history type for spring Chinook salmon. The 3000 trajectories used for Chinook means that there would be 30 trajectories or fewer for some life history types. For steelhead, where some life-history types have a proportion of 0.02%, the use of 2000 trajectories results in less than 1 trajectory for that life history type. Confirming convergence is a typical numerical check done with simulation of individuals. In a comparison of the model predictions with increasing numbers of individuals, the predicted values should reach an asymptote once convergence is attained. Such convergence means that the predictions are not dependent on the number of individuals trajectories simulated and therefore representative of the internal dynamics of the model rather than how many individuals are followed.

Equilibrium Abundance (N_{eq}) defined as where BH curve intersects $X=Y$ line. Diversity is defined as the proportion of trajectories that are used to compute equilibrium abundance. The model

estimates capacity and intrinsic productivity for each trajectory, from which the BH model equilibrium can be directly calculated (for a deterministic model). The report should clearly explain this or provide the equations. If one conditions only on trajectories that persist, does this bias the estimates of equilibrium abundance? Decisions on how extinction during simulations affect calculation of output variables should be documented.

The EDT report would be improved by explaining how the proportions of the population in different ages in freshwater and the ocean were determined for spring Chinook and steelhead (Table 2-2 and 2-3). The EDT modeling team explained that they derived overall age proportions for spring Chinook from ODFW (2020), but that document only reports total age, not freshwater and ocean age breakdowns. They indicate that they used "professional judgement" for the finer breakdowns, without a discussion of the uncertainties. For steelhead, the EDT team said they used data from Johnson Creek, but that is an ecologically dissimilar system not directly comparable to the rivers assessed in the report.

No measure of uncertainty is illustrated in the bar charts. Only three years are used to represent range of conditions with one water-year for each condition. The analysis would be more powerful if several years were used for each of the dry/normal/wet year categories to determine variation within water year types and variation among water year types.

The discussion of productivity being higher in dry years gives some insight into how the EDT models work: in dry years, less habitat is available, which reduces capacity, but that habitat is of higher average quality resulting in better productivity. Because equilibrium abundance is lowest in dry years, this suggests that the capacity effect is dominant. It would be interesting to know which life stages are driving this effect.

The EDT modeling team explained to the ISAB that the "Avg" bar is just the average of the three dry/normal/wet values (Figure 3-1 in ICF 2022). These three categories could also be weighted in an averaging to reflect the proportion of water years. Projected trends in climate could be incorporated in the analysis.

Only about 40% of trajectories lead to a "sustainable" system where the number of fish returning is >1 for each fish that started (Figure 3.2 in ICF 2022). If the modelers could determine which reaches/life histories are "sustainable" based on where the trajectories originated, it could help managers determine what type of remediation is needed for reaches that are not sustainable.

Figures 3-5 to 3-8 provide a nice cross-comparison across alternatives, but without any measure of prediction uncertainty, it is difficult to simply interpret whether these differences are meaningful. These figures show the percent change from NAA, which is helpful. A major missing

source of uncertainty is due to use of a single “representative” water-year for each water year type.

Diversity reported by the EDT models is not equivalent to species diversity or life history diversity in most ecological studies. It represents the diversity of habitats that have trajectories that are sustainable. The report could clarify this because readers may assume that EDT-predicted diversity represents life history diversity.

ii. Are the key functional forms (e.g., shapes, dependence on hydrological and environmental variables) sufficiently realistic?

The EDT models do not use specific functional relationships for major process that influence populations of Chinook salmon and steelhead. Rather, they reduce abundances based on rules and indices for potential relative influence. Rule-based interpretations of projected conditions under a set of management alternatives are used to reduce abundances for each life stage and stream reach. Most of the relationships are derived from literature from different populations and their relevance to the WVS should be assessed.

Several key factors are fixed throughout all analyses. Life history parameters are fixed and do not change under different flow or environmental conditions. Marine survival is entered as fixed survival rates. Trajectory sets and river channel dimensions are fixed for the alternatives to reduce variation due to different trajectories and/or system geometries.

The responses to hydrologic flow variation are based on indices rather than explicit functional relationships. The historical and current flow patterns are compared to determine whether projected flows are different from historical patterns to assess the differences in inter-annual flow variability at high and low flows (Lestelle 2005). An index of 2 is assigned if there is no change, and higher and lower indices indicate shifts toward more or less interannual variability and decreased or increased low flows. These are general influences of hydrology rather than explicit functional relationships between survival and flow. While such use of indices is reasonable, it does hinder easy determination of why the EDT models predicted certain responses.

Differences in physical habitat (width and depth of reaches; spatial extent of reservoirs, river depth and velocity) between alternatives was not explored in the EDT models. Floodplain inundation and differences in flood refuge and winter habitat are not included in the models.

iii. What aspects of the model are stochastic?

Except for the sampling of the model domain with fish trajectories, the EDT models are deterministic. The EDT analysis for the WVS PEIS evaluated 7 alternatives x 3 water years (wet =

2011; dry = 2015; normal = 2016) for each of spring Chinook and winter steelhead. Within-year stochasticity of environmental conditions is not included because three specific water-years are evaluated. Several water years for each wet/dry/normal would be needed to see how consistent the results are over variation in environmental conditions.

- b. Are the models formulated to assess fish responses at the appropriate temporal and spatial scales that both capture the effects of the alternatives and generate outcomes relevant to management scales?*

The temporal scale is monthly time step. The spatial scale encompasses reaches in the Santiam and McKenzie rivers. The mainstem Willamette River flow, floodplain inundation, and habitat area are not included in the models. These scales are appropriate for the purpose of the model of assessing the habitat effects included in the models. Effects of the dams and changes on downstream flow and water quality are not represented in the models, which limits the inclusion of potential effects of the dams and flow management on life stages while they reside in or migrate through the mainstem.

- c. Are the estimates used for model input parameters (hydrological, environmental, processes) reasonable and scientifically defensible? What data are used to estimate or confirm model parameter values?*

Habitat conditions (p. 2-9) are imported from existing models for all but the Santiam River. A new EDT model was built for the Santiam, and more thorough description of Santiam would be useful. Habitat conditions were obtained from several sources, and further information on the consistency of the habitat values and representation of the effects in different portions of the subbasins would be helpful.

For each alternative, temperatures throughout the river come from the CE-QUAL-WE model. Res-Sim provides information on flows. Fish passage comes from FBW. Outplanting and gravel augmentation is included for all scenarios (this increases small-riffle by 10%, which causes other types to decline). Total dissolved gas model results are not included in this analysis.

- d. What sources of variation that affect population dynamics and responses have been accounted for?*

The EDT models are predominantly deterministic. Life-history variation is based on the multiple runs for individual trajectories. As mentioned above, life history parameters and several key factors are fixed throughout all analyses and do not change under different flow or environmental conditions. Marine survival is represented as fixed survival rates. As noted above, trajectory sets and river channel dimensions are fixed for the alternatives to reduce variation due to different trajectories and/or system geometries.

- e. *Have the researchers adequately evaluated the sensitivity of the predictions to uncertainties in various sources of data?*

The EDT report does not evaluate the sensitivity of the models to uncertainties in the multiple sources of habitat information or life history parameters. A study in the Yakima basin reported that EDT was relatively insensitive to flow attributes (reported in Steel et al. 2009). The EDT models do not directly represent flow variation and uses output from FBW to represent the influences of dam operations on passage.

- f. *How sensitive are predictions to uncertainties in hydrological conditions, environmental variation, and model parameters?*

Sensitivity analysis was not used in the interpretation of the results.

- g. *Are subcomponent models adequately integrated into the larger model?*

There is no evaluation of whether the rules for different aspects of survival and movement collectively produce realistic overall population-level predictions.

2. Key assumptions and limitations

- a. *Are the model assumptions and limitations sufficiently documented? What are the key assumptions, strengths, and limitations?*

The EDT approach is complex, and the EDT Report (ICF 2022) assumes that the reader is familiar with the EDT methodology. More thorough description of the EDT modeling approach would strengthen the use of the models' results in the PEIS consultation process.

The EDT models were purposely designed to focus on habitat effects and so can only provide partial answers to problems that require a full life approach. Thus, EDT is best used to address questions about habitat or are used in combination with other models to express responses integrated over the full life cycle.

A limitation to the WVS PEIS application is that the EDT models assess the consequences of flow management on habitat using single years to represent wet, normal, and dry years. As we move toward warmer and drier years, the low flow conditions may become the new average and demand stricter adherence to biologically meaningful minimal flows.

The EDT models assume that the projected changes in temperature, flow, fish passage, and such reflect habitat quantity/quality differences between current conditions and alternative scenarios under different water year conditions. The EDT report is clear that it represents a short-term response to the habitat conditions for a year.

There are many life-history assumptions, especially for steelhead. As in other models in this review, the steelhead model does not include interactions with resident rainbow. The steelhead models relied more on information from outside the basin because of a lack of steelhead data for the basin.

Effects of hatchery fish on natural-origin fish are not included in the EDT analyses.

- b. Is the reasoning, rationale, and evidence for the key simplifying assumptions (e.g., no interactions with other species) reasonable and adequately documented?*

The relationships and assumptions for the rules and indices are not thoroughly explained in the report. The report cites earlier reports (Lestelle 2004, 2005), but these relationships are not discussed relative to the Willamette basin and it is not clear how much of the information in the models can be confirmed as representative of the WVS.

The EDT steelhead model assumes complete ecological and reproductive separation between resident rainbow trout and steelhead. This is known to be untrue, so the real question is whether it is important. In many populations, there is considerable gene flow between forms, and phenotypic plasticity affects which life history pattern the fish display. There may not be sufficient data to specifically model this at present but at least some acknowledgement of the fallacy of the assumption and implications should be included.

3. Model validation and verification

- a. Does the model documentation adequately describe testing steps utilized during model development (i.e., consistency check, sensitivity analyses, calibration, validation)?*

The report does not mention or discuss calibration, validation, verification, or sensitivity. The models are combinations of many stage and habitat-specific submodels, all with the Beverton-Holt form, each derived by a complex rule set based on a combination of actual data and expert opinion. There is no evidence presented that the combination of these submodels produces realistic results at the population level. Most of the relationships are drawn from literature values for other unspecified systems. More thorough description of the validity of these relationships and presentation of reality checks on when the submodels are combined together would allow for assessment of the appropriate confidence level to assign to model predictions.

- b. Have the researchers adequately assessed the fit of model to current data? Have the researchers assessed the fit of the model to new data (e.g., use it to predict years not used in fitting and compare)?*

The analysis of EIS alternatives used three specific years to represent the range of flow conditions. No assessment of model fit specific to those three years was reported.

- c. *Has the model programming or system components been tested for computational correctness and associated errors? If not, what is the potential for errors to occur?*

While no code verification, QA/QC or testing is described, the EDT model has been used for several decades and was used previously in the Willamette basin. There always is potential for computational errors in new applications. The EDT report does not indicate how code verification and QA/QC was performed.

- d. *Has sensitivity and uncertainty analyses been applied to the model and the results used effectively to identify sensitive parameters and quantify variability around model predictions?*

As indicated previously, the EDT report does not evaluate the sensitivity of the models to input uncertainties.

4. Data gaps

- a. *For which part of the model is data strongest and weakest? Is the model sensitive to these data gaps?*

The models represent habitat conditions in the tributaries, and abundance of juvenile salmon and steelhead are modified according to rules for habitat relationships in the models. The models are scientifically sound when used for their intended purpose of habitat assessment. The lack of steelhead data specific to the Willamette basin leads to relatively high uncertainty in the WVS-specific predictions for steelhead. Also, the previous comments about the use of data from other systems and rules developed for the Columbia River basin are relevant. The gaps are difficult to identify because of the general form of the rule-based approach. Obtaining site-specific information for the critical inputs used to derive the rules would reduce model uncertainty.

- b. *If so, then what type of research program is needed?*

Development of life history parameters and rules that are based more from the Willamette River basin would provide better inputs for this and other models of Willamette salmonids.

5. Model output

- a. *Can the outputs of the model be used to rigorously compare and contrast different management actions?*

The models can be used to compare and contrast the responses to the alternative management actions. The confidence level of using model predictions in absolute terms is unknown. The models are well suited for ranking alternative management actions and using the model

predictions as relative changes. Without site-specific model-data comparisons, model predictions interpreted as absolute abundance may have biases and high uncertainty that can hinder distinguishing responses among the alternatives.

- i. How reliable are the comparisons of the effects of the alternatives on spring Chinook salmon and winter steelhead (e.g., can be used only to rank alternatives; can be used in cost-benefit analyses)?*

See response to 5g above.

- ii. Is the output stochastic so questions about quasi-extinction can be answered?*

Because EDT is not a complete life-cycle model, it cannot be used to address questions of extinction risk.

- iii. Do the researchers distinguish whether differences in predicted responses among alternatives are biologically meaningful?*

No, but they provide the results to the Corps for such analyses.

- b. What specific metrics or types of output are provided to the user? Are these outputs sufficient for evaluating the EIS alternatives? Would additional types of output be useful?*

Output metrics are identified in Table 1. They are not in themselves intended to be sufficient for fully evaluating EIS alternatives, and they provide relevant information to the evaluation.

- c. How far outside of the envelope of data used to develop the model will climate change take us? Do we have good information on how the model will perform outside of its development based on historical conditions?*

The performance of the models under climate change is unknown. It would be necessary to see if the range of data supported by the EDT models encompasses likely scenarios under climate change.

6. Model improvement

- a. What improvements can be made to address any critical shortcomings of the models?*
- b. Where should future model development be focused?*

More attention to validating the model outputs (checking against data) would be very helpful. An assessment of sensitivity of results to assumptions and inputs for this particular application would improve confidence in the results. Analysis of sensitivity and uncertainty is needed.

4. Editorial comments

The description of the EIS alternatives and how they were implemented in the analysis is clearly presented. Also, the reporting of results by tributary and collectively for the areas modeled is informative.

A reader of this report likely will not understand how the EDT models work. A summary document showing an example of a trajectory would help the reader understand how the modeling process works. The bibliography does not include any reference for the beginner for background information. The Blair et al. (2009) paper is better than the references provided, but it still is a complex description of the modeling approach.

Descriptions and references for the more recent versions of EDT are not provided. The EDT modeling team provided the following information on the different EDT models to the ISAB, and such information should be included in the report:

EDT3 supports backwards compatibility with EDT 2—that, EDT3 can produce the same outputs from the same inputs as EDT2. However, the default settings for EDT3 are different than EDT 2. Backwards compatibility is preserved by supporting the EDT2 version of each setting listed below as a user-configurable option.

EDT 3 adds the ability to calculate survival factors independently for different habitat types in the same stream reach; this capability is used for modelling differences between littoral and limnetic habitats in lakes/reservoirs as well as inundated floodplains.

EDT 3 simplifies the size classification scheme for stream reaches. EDT 2 classifications based on maximum width and maximum flow in different calculation contexts; EDT 3 consistently uses maximum flow.

Page 3-1 of the ICF Report. First paragraph uses Table 3-1 twice (one should be Table 3-2). Similarly, there are two references to Figure 3-1.

The long appendices appear to have details on how certain parts of the EDT were modeled, but there are inconsistencies. For example, Appendix B states that “There are 15 EDT reaches within this stretch of the mainstem of Willamette River, starting at Willamette Falls and ending at the split between the Coastal and Middle Fork Willamette (Appendix B-1).” which results in 16 graphs in Appendix B-1? Why the discrepancy between 15 and 16? Similarly, the text says

there are five reaches in Middle Fork, but more than 5 graphs presented in Appendix B-2. (Text on page A-3 omitted the number of reaches in Fall Creek)

E. Review of OSU Fish Survival and Flow Models

1. Summary Comments

The ISAB finds that, in general, the approach is reasonable and well supported for relative comparisons of alternative flow scenarios for the below-dam portions of the WVS, and that the OSU survival and flow models are generally well-constructed cohort models for portions of the Chinook and steelhead life cycles. Because these models cover only portions of the life cycles and are not integrated across years, they are not appropriate nor intended for evaluating future population-level effects and should not be used by themselves to assess ESA-related risks.

Strengths

- The use of Structured Decision Making as an overall context for developing and implementing the models for decision-support for flow management is laudable because it incorporates multiple perspectives into a clear and repeatable structure for model development, application, and revision.
- Using a formal optimization algorithm is an excellent and reliable way to identify beneficial flow patterns when there is high confidence in assumptions.
- Details of the numerous submodels were described well, allowing a full evaluation of functional forms and data sources.
- The inclusion of sensitivity analyses provided a context for evaluating the importance of potential errors in different submodels.
- The models' spatial domain from tributary dams down to Willamette Falls provides explicit treatment of flow, habitat, and temperature interactions for this portion of the basin that are not included in the other models reviewed in this report.

Limitations

- The bottom-up approach to building the cohort models from numerous process submodels without a full calibration of the results, along with the uncertain functional forms and limited data sources for some submodels (detailed under Question 1.a below), makes evaluation of the reliability of the overall models difficult.
- A key limitation for the application evaluating the WVS PEIS alternatives (Peterson 2022b) is that it was done for only three pre-selected years (2011, 2015, 2016) to demonstrate survival outcomes from a representative range of flow conditions,

although the Corps did not use the OSU models for this purpose in the Draft EIS. It is not known whether rankings of alternatives would be different for a broader selection of years, whether flow variability among years might affect survival, or if these years realistically reflect likely future conditions in the WVS.

- The limited spatial domain, excluding the river below Willamette Falls and tributaries above dams, place certain limits the predictions of the models that should be factored into the interpretation of responses across alternatives. Excluding tributary flows could hinder estimating juvenile and outmigrant responses to flow and summer temperatures for returning adults.

Recommendations

- The ISAB's review of the OSU survival and flow models relied on two papers published in scientific journals and a supplemental document. Reliance on journal publications carries both benefits and challenges. The primary benefit is that the published models and papers already received scrutiny and rigorous peer-review. However, this is also a challenge because as issues arose and were addressed (often minor and clarifying) through correspondence between the authors and the ISAB, the explanations are not incorporated into the publication and thus not fully available to wider audiences. The ISAB recommends that when these models are used in the future, a single document should be prepared that fully documents the model structure, data sources, statistical fitting, sensitivity analyses, and applications in one place.

2. Background: OSU fish survival and flow models

The OSU fish survival and flow models were developed in the context of a multidisciplinary team of scientists, stakeholders, and managers known as the Science of Willamette Instream Flows Team (SWIFT). The models were developed as a decision-support tool to characterize annual water management tradeoffs, including total dissolved gas, flow, and habitat availability (Chapter 3 in the DPEIS). Two published journal articles described the application of the models to the WVS. DeWeber and Peterson (2020) described the general approach and its initial application to WVS salmonids. Peterson et al. (2021) described the details of models that were revised following stakeholder review of the initial application and presented further results of the analysis.

The OSU fish survival and flow analysis used Structured Decision Making (SDM; e.g., Conroy et al. 2008) as a framework for linking river flows resulting from management actions with

ecosystem or social benefits. As implemented here, SDM consists of six steps (DeWeber and Peterson 2020):

1. Identify the decision context and objectives (here, specifically, selecting an alternative that is consistent with the Congressionally authorized purposes of the dams, and poses “no jeopardy” for spring Chinook and winter steelhead as identified in the ESA recovery plans for Upper Willamette River ESUs).
2. Identify the management alternatives.
3. Break down the problem into components and build a model based on scientific knowledge (here, flow-related and passage impacts on salmon and steelhead survival).
4. Use the model to predict and compare results of alternative management actions.
5. Evaluate the model’s sensitivity to scientific assumptions.
6. Implement the best alternative.

DeWeber and Peterson (2020) described a pilot phase application of the SDM framework as implemented by SWIFT to assess ecological effects of flow management for six different fundamental objectives (DeWeber and Peterson, Table 1). These objectives were refocused into six means objectives (i.e., objectives focused on the means of achieving fundamental objectives; Gregory et al. 2012), along with six corresponding metrics to be predicted and evaluated. The authors then established five different flow scenarios that specified seasonal minimal flows for the mainstem Willamette River at Salem and evaluated the metrics for each objective under each scenario using simple data-driven models (that is, models based on data series of about 20 years). Ultimately, they found little difference in results among the scenarios. The authors interpreted the similarity in results among scenarios as being mainly due to “partial control” of the hydrosystem. Specifically, flows varied but were generally greater than the minima specified in the scenarios, even in below-average water years. They concluded that the SDM framework was useful for structuring information and values to support decision making, but a further iteration of the SDM process was needed to refine the decision framework to include tributary flows and to refine objectives and models.

Peterson et al. (2021) described the quantitative decision-support models developed during this second iteration (phase 2). Rather than including all the objectives identified by DeWeber and Peterson (2020), these models restricted their focus to flow effects on spring Chinook salmon and steelhead trout. They developed four decision-support models reflecting adult and juvenile life stages for each species. The spatial domain of the models extends from Willamette Falls Dam up to the WVS dams in the main salmon-bearing tributaries (North and South Santiam, McKenzie, and Middle Fork Willamette rivers). The models’ outputs then helped

identify optimal flow regimes for below-average flow years, from which managers developed candidate management strategies that were subsequently evaluated with the models. The four models identified optimal flow regimes that (1) differed from existing regulations, and (2) varied substantially among species and life stages, reflecting tradeoffs among the objectives evaluated. Despite these tradeoffs, optimizing across all objectives with equal weighting suggested a flow regime with only small losses for most objectives compared to the individual optimal solutions. The authors also identified water temperature as the primary driver of the outcomes and suggested that managers explore means of controlling for biologically optimal temperature ranges as well as flows.

Peterson (2022b) went beyond the two published papers to describe results of their analyses to support the WVS PEIS process. The report used the four existing decision-support models (Peterson et al. 2021) to assess salmonid responses to the EIS alternative flow regimes. Model structures were the same as those in the earlier paper, but the models' inputs included revised juvenile Chinook habitat definitions, flows, and water temperatures specific to the PEIS alternatives. The Corps provided estimates of discharge and water temperature for seven flow management alternatives for three representative years: 2011, 2015, and 2016. These years were selected as representing the expected full ranges of precipitation and temperature via a collaborative process among USGS, SWIFT, and Corps, and are the only years for which the modeled stream temperature datasets were generated. Streamflows were estimated using the USACE Hydrologic Engineering Center (HEC) ResSim model (Klipsch et al. 2021) and temperatures were modeled using the Portland State University Water Quality Research Group CE-QUAL-W2 hydrodynamic and water quality model (Wells 2020). Results were consistent with the earlier analysis, with no single alternative being best for all four outcomes and relative effectiveness of the alternatives differing among species, life stages, and years.

The ISAB recognizes that tracking how these models are used in the EIS and future management will prove challenging to interested stakeholders. Specifically, model descriptions and parameters are distributed throughout several documents (two published papers, a technical supplement with submodel and parameters, two responses to ISAB questions that included clarifications and corrections for some submodels, and a brief final report with results for the EIS alternatives). Moreover, the various documents differ in the years simulated and other key aspects. To facilitate broader clarity, the documentation of the analyses would benefit from a single source having all the information or at least a summary with a "roadmap" to show how the constituent documents are connected.

3. ISAB Answers to Review Questions

It should be noted that SWIFT developed an analytic framework that included effects on Chinook salmon, steelhead, black cottonwood, and herpetofauna (DeWeber and Peterson 2020), but ISAB comments focus only on the salmon and steelhead analyses. Moreover, at the time of the ISAB's review of the OSU models, the ISAB did not know the details of how this set of analyses would contribute to the multi-model approach used in the WVS PEIS. Thus, this review treats the OSU models for the WVS PEIS as if they were standalone analyses and not part of multi-model approach. Without knowing how the results will be ultimately used, the review is limited to general aspects of the analyses.

1. Model structure and development

- a. *Is the model structure appropriate for evaluating fish responses for the EIS alternatives?*
 - i. *Are the relevant processes (mortality, growth, reproduction, movement) included and are they represented in a way that encompasses the changes within and among the alternatives?*

The OSU fish survival and flow models partially address how Chinook and steelhead will respond to the WVS PEIS alternatives. First, the ISAB identified the models' strengths: (1) they have been designed to enable optimization of flows; (2) they are well grounded in approach as cohort models; and (3) they are informative with respect to the interpretation of results. Second, these models present challenges because: (1) confidence ranges for their outputs are either unknown or unreported, and other sources of uncertainty (such as submodel structure and the applicability of data sets not specific to the modeled populations) are not discussed; (2) they report and use dimensionless outputs and metrics that are then averaged, resulting in difficulties in relating summarized results with *in situ* conditions; (3) they do not include effects across generations; (4) by using only three "representative" years, they do not reflect full variation of climate conditions; and (5) they do not include a description of data sufficiency or adequacy for the models.

Petersen et al. (2021) described the modeling approach. There are four decision-support models: one each for juvenile and adult stages of Chinook salmon and steelhead. These were designed to generate information for the SDM and are purposely not full life-cycle models. Thus, they do not account for all ecological processes (or conditions) and potential impacts of flow management experienced by the fish. Reproduction and growth are included in the juvenile Chinook and the adult steelhead models. The authors noted that streamflow within a year affects adults and juveniles from different cohorts and indicated that this complicated the modeling and "prevented the use of traditional life-cycle models."

Each model uses temperatures and discharges for individual years to predict fish life stages through the various submodels to determine abundance of the endpoint life stage. The adult Chinook model estimates the number of redds for which fry survive to emergence, and the juvenile model estimates the number of smolts arriving at Willamette Falls expressed as adult equivalents (i.e., the expected number of adults returning to the river, calculated using estimated ocean survival based on tag recaptures for 1999-2018). The adult steelhead model estimates the number of age-1 juveniles, and the juvenile model estimates the number of smolts surviving to Willamette Falls. The outputs from each model are converted to utility values, which are then summed across all models. The utility values are used in two ways: in an optimization procedure to identify optimal flow regimes (Peterson et al. 2021), and to compare WVS EIS alternatives (Peterson 2022b).

Each of the four models has several submodels described in the supplementary materials (Peterson et al. 2021, Suppl. S1-S5, with corrections in Peterson 2022a). For example, the Chinook adult decision-support model has submodels for dam passage, *en route* survival, migration proportions to tributaries, prespawn survival, proportion of hatchery origin, redd dewatering, and such to take an input number of returning adult Chinook at Willamette Falls and output number of redds surviving to emergence by tributary. Submodels are also described for below-dam temperature (Peterson et al. 2021, Suppl. S6) and river discharge (Peterson et al. 2021, Suppl. S7 and S8). A few variables affect each submodel; for example, the variables that drive the Chinook adult model are day of the year, average daily discharge, average daily temperature (above a cutoff), indicator variables for tributaries, year-to-year variation modeled as random effects. Peterson et al. (2021) states: “During a given water year, streamflows affect returning adults and juvenile salmonids from previous brood years. This prevented the use of traditional life cycle models to evaluate candidate flow regimes.” The reasoning for configuring the models as was done (and not full life cycle) should be clearly explained. Was it for ease of bookkeeping or because the models were scaled for generating the predictions that were needed? There is a good rationale for the modeling that was done and should be further explained.

In conclusion, the answer to the question posed is yes: the relevant processes are included and can be used to evaluate the different regimes to the extent that the endpoints of the juvenile and adult cohort models are relevant. Of the seven EIS metrics (Table 1), these methods only address the last two (survival).

- ii. Are the key functional forms (e.g., shapes, dependence on hydrological and environmental variables) sufficiently realistic?*

In general, the functional forms (submodels) are well reasoned and supported. As with all models, which by definition must be simplifications of the real system, there are comments

about the functional form and process included. The ISAB presents their comments as general and then specific for each of the models.

First, it is difficult to evaluate the estimation of the submodel parameters and how those submodels fit together into the overall models. The authors calibrated or fit each submodel (stage-process combinations) separately, which is a sound approach. However, it is not clear that all of the submodels fit together to generate realistic population-level dynamics. The results may be presented and available, but the multiple documents make it difficult for the reader to identify the relevant model-data comparisons and then determine how those comparisons were performed. For example, Peterson et al. (2021) used 2000-2018 data, but the draft EIS analysis (Peterson 2022b) focuses on only three of these years.

Second, the ISAB has concerns regarding extreme parameter estimates in several of the logit-scale regression models, where coefficients are very high or low. Any logit estimates that are less than -20 or greater than +20 should be considered as very unstable, i.e., small changes in data lead to large changes in estimates (Rindskopf 2002). To alleviate biases created from these extremes, it may be better to rescale the data so that estimates fall within +/- 20. As an example, for the juvenile Chinook passage survival submodel this could be done by subtracting 25° C from every temperature so that the intercept is the survival at 25° C.

Chinook adult model (part of 1.a.ii)

- For adult en route survival, the estimated intercept (62; Peterson 2022a) is far into the realm of numerical instability (see general comment above). This regression should be rescaled to fix that problem.
- The redd dewatering ratio was not explicitly defined in the initial papers, and the response to the ISAB's request for clarification provided an adequate definition. This should be added to the document.
- In the initial paper (Peterson et al. 2022, Supplement), egg hatch temperature threshold has a value of 1028, with undefined units. This was clarified in subsequent responses as being degree-days, although the authors did not specify the time-period or temperature basis for calculating degree-days.
- What is sometimes termed *prespawning mortality* (PSM) is here divided into two components: *en route* survival, which is temperature-dependent, and prespawn survival which is calibrated to spawner survey data with a complex model that includes subbasin, accumulated degree days from passage at Willamette Falls, proportion of hatchery origin fish, and run size. Regarding *en route* survival, Peterson et al. (2021) reported erroneous

parameters for the temperature relationship based on information published in Sullivan et al. (2000, section 4), but provided the ISAB with corrected values (Peterson 2022a). However, the analyses reported in Sullivan et al. (2000 App. C) for Chinook salmon all are for juveniles or jacks, not for full adults, and the citations for the jack survival data appear to be extraneous; the relevance of the data used should be further documented.

- For prespawn survival, the submodel includes the proportion of hatchery origin fish, but no information is provided about harvest rates on these fish. Harvest of these fish may be a built-in component of the submodel but warrants a comment on its influence. It has recently been reported (Bowerman et al. 2021) that prespawn mortality is size-selective, being more severe for larger females. Thus, a size-selective model may warrant further exploration, both here and for the other Willamette models.
- In the redd-survival submodel, other than temperature, there is a fixed carrying capacity for redds, such that each female coming after that cap is reached displaces the entire egg production of a previous female, resulting in no additional surviving embryos. This neglects the shift in female body size that is typical of salmonids (larger = earlier; smaller = later), possible expansion of habitat used as density increases, and other features of salmon ecology. Moreover, it is not clear whether the ecological complexities of stream life and the duration of redd defense are included in the model. This is also an issue for the other Willamette models.

Juvenile Chinook model (part of 1.a.ii)

- In the juvenile passage survival submodel, estimates (logit scale) should be recomputed after appropriate scaling to avoid (logit) estimates <-20 or >+20.
- The models assume that juveniles not finding suitable habitat move downstream to the nearest suitable reach. However, salmonids commonly move upstream, including into non-natal tributaries (e.g., Murray and Rosenau 1989; Scrivener et al. 1994; Bradford et al. 2001; Daum and Flannery 2011; Phillis et al. 2018), so this should at least be acknowledged, even though there may not be data available on this for the system in question.
- Temperature is treated as the only factor driving growth, but food availability is often also important (see recent review by Railsback 2022). The authors recognize this issue (Peterson 2022a) but have no data to support a model incorporating food supply; they relied on expert opinion to suggest assuming 2/3 of full food ration in the model. This could be a significant assumption and key uncertainty of the model in some situations.

- The supplemental Table S2 of Peterson et al. (2021) reports a fecundity of 4950 from Healey (1991), but the ISAB could not find this specific value in Healey's chapter. The ISAB requested a clarification, which was provided by Peterson (2022a) and noted that the citation should have been to Healey and Heard (1984), averaging fecundity estimates for Columbia River samples and rounding upward to 4950. While there is no discussion of how accurate this rounded estimate is for Willamette River Chinook, the value will only affect absolute abundances and should have little or no effect on relative comparisons of model predictions.
- The model uses a single average fecundity value, with standard deviation to represent uncertainty in the value. While this approach, like that of the other models reviewed here, does not adequately represent the variation contributing to egg production resulting from body size and population differences, these details would only affect the absolute numbers of juveniles produced and are unlikely to affect relative comparisons across scenarios.
- In Peterson et al. (2021, Supplement S2), survival from egg deposition to emergence is 0.68 for temperatures below 16° C and 0.34 for temperatures above 16° C. This does not account for other sources of variation in survival from egg deposition to emergence related to flows known from other studies (e.g., Anderson and Topping 2018) or other sources of mortality, such as fine sediment. It is unclear whether these estimates are realistic, either in terms of survival or the prevalence of these regimes. For example, Murray and McPhail (1988) reported survival from fertilization to emergence of Chinook salmon as 46% at 14° C (their highest temperature) vs. 87% at 11° C, 90% at 8° C, etc.

Steelhead adult model (part of 1.a.ii)

- Adult *en route* survival, probability of spawning, and juvenile rearing survival estimates on the logit scale are again in the region of numerical instability and some rescaling is recommended – see general comment above.
- The issues raised regarding *en route* and prespawn survival for the adult Chinook model also apply here. Also, Cramer (2001) is cited as the source for the prespawn survival temperature relationship, but we find no data for steelhead adult survival in that report.
- The fecundity estimate (2912 eggs/female) cites Bulkley's (1967) study on the Alsea River. Peterson (2022a) noted that the values in Bulkley were modified for smaller fish based on a more recent study (Firman et al. 2003). That study reported a relationship between length and fecundity, and an average of 3438 in the sampled fish. There is now a great deal more data on fecundity in steelhead and Chinook salmon (e.g., Quinn 2018 and data associated

with this review), and so this variable could use more explanation and justification, and perhaps be adjusted for accuracy.

Steelhead juvenile model (part of i.a.ii)

- Some of the logit-scale models result in extreme estimates; rescaling is recommended.

iii. What aspects of the model are stochastic?

The four cohort models are probabilistic in that they assign fish to tributaries and stream segments (upstream migration) or to stream segments by size-class (downstream migration) based on binomial distribution models. The models select from a set of water patterns plus random start point from a fixed distribution and then calculate the output. This is repeated multiple times.

The modelers attempt to induce stochasticity within the submodel linkages by using binomial or multinomial distributions to move between stages. However, the number of adults or juveniles is sufficiently large that the binomial stochasticity should be negligible at the whole-population scale, though perhaps not for individual juvenile stages and spatial subdivisions. As well, there is likely to be “autocorrelation” in the submodel steps, i.e., a year with poor weather conditions could make all transitions fall below the “mean” simultaneously, which is not captured. This would imply that, overall, the reported variability in output for a given set of inputs underestimates the true uncertainty in the results.

b. Are the models formulated to assess fish responses at the appropriate temporal and spatial scales that both capture the effects of the alternatives and generate outcomes relevant to management scales?

The temporal scale is seasonal (varying duration by model) with a weekly time step. The spatial scale of each model extends from Willamette Falls to tributary dams, with the whole river domain divided into 18 reaches. This is adequate for the cohort models as formulated, with the recognition that it excludes much of the life history of the species, as discussed under Question 2.a below.

c. Are the estimates used for model input parameters (hydrological, environmental, processes) reasonable and scientifically defensible? What data are used to estimate or confirm model parameter values?

The models use a generic logistic regression approach to estimate most of the parameters. While this is a widely used technique, it is not clear that the logit-linear relationship is an appropriate form for the various processes modeled here, and there is no consideration of alternative regression models. Some of the data comes from literature from different

populations and is not necessarily applicable to these populations. This is a difficulty in any effort to build a bottom-up model of cohort dynamics, but some consideration should be given to the errors that may result.

For the logistic regression submodels, the modelers need to be very careful when estimates of parameters tend to +/- infinity, otherwise many numerical issues present themselves (Rindskopf 2002). For example, in the steelhead adult model, the two alternative probability-of-spawning submodels have intercepts of -25.6 and -30.0. In cases like this, the day variable should likely be rescaled as starting in mid-year rather than from the start of the calendar year.

d. What sources of variation that affect population dynamics and responses have been accounted for?

To reframe the question “are there sources NOT accounted for?” We identify a few above, including effects of female body size on fecundity, potential habitat expansion as capacity is reached, potential up-river movement of juveniles.

e. Have the researchers adequately evaluated the sensitivity of the predictions to uncertainties in various sources of data?

Peterson et al. (2021) performed two different sensitivity analyses. One analysis identified which components had the greatest influence on estimated outcomes for the best-performing flow regime. Another analysis evaluated the sensitivity of rankings of flow regimes to model components and to evaluate the expected value of perfect information as a measure of the potential benefits of improving information for various components. The methods are promising and should be more fully described to ensure proper interpretation by others.

f. How sensitive are predictions to uncertainties in hydrological conditions, environmental variation, and model parameters?

Results of the sensitivity analyses suggest that the outcomes are highly sensitive to temperature via the temperature-dependent survival submodels at various life stages, with other substantial sensitivities to initial numbers of fish and submodels related to habitat capacity (Peterson et al. 2021). Sensitivity to parameter uncertainty was not directly addressed.

The ISAB highlights the potential effects of error propagation, in that each time a quantitative submodel output is used, any small inaccuracy or error can compound to bias the performance or output for the “overall” model. Because each cohort model is fed by multiple submodels, such potential compounding of error needs to be addressed for importance and effect on overall results.

The authors dealt with key uncertainties in several ways. They used stochastic realizations, repeated analyses with three different assumptions about habitat suitability estimates and performed a one-factor-at-a time sensitivity analysis. How the sensitivity analysis was used was not documented. In general, few diagnostics were reported to explain the few situations of output-scenario combinations that showed relatively large responses.

g. Are subcomponent models adequately integrated into the larger model?

The OSU models used standard methods to integrate subcomponents into the four cohort models. This analysis is not used in isolation but rather will be integrated with other population-level analyses.

2. Key assumptions and limitations

a. Are the model assumptions and limitations sufficiently documented? What are the key assumptions, strengths, and limitations?

Neither paper (DeWeber and Peterson 2020; Peterson et al. 2021) has a section or table explicitly identifying critical assumptions, potential sources of bias, confidence that assumptions are valid, and risks to salmon and steelhead resulting from potential errors. A number of these are identified throughout the manuscripts, but it would be helpful to have them collected and presented as a table or text description in the final EIS documentation for this and other models. While there is generally good recognition of key assumptions in the decision model component descriptions, there are no sensitivity tests for most of these, so there is little basis for judging their importance. Below, we discuss some of the notable assumptions and limitations of this analysis.

A major limitation is that the models are fitted to data from relatively few years, which vary among submodels and then used to predict different time periods (2000 to 2018 in Peterson et al. 2021; three selected years in the EIS analysis). As we move toward warmer and drier years, low flow/warm water conditions may become the new average and demand stricter adherence to biologically meaningful minimal flows. This is related to problems of predicting outside the range of the data space in regular regression. Further sensitivity analysis regarding this issue would lead to more confidence in the results. Ultimately, this is not a criticism of what was done, but rather an issue that needs thought and transparency for the EIS and selected alternative.

DeWeber and Peterson (2020) identify a “partial controllability” problem due to observed flows rarely reaching minima that differentiate the different flow patterns. This appears to limit their flow optimization analysis and could limit comparisons of the PEIS flow management alternatives.

While including river reaches of the Willamette not included in the other models is a strength of this model, the limited spatial domain (i.e., excluding the river below Willamette Falls and tributaries above dams) may be a significant limitation for other factors. Excluding tributary flows could be a substantial problem, especially for estimating juvenile and outmigrant responses to flow, and summer temperatures for returning adults.

There are many life-history assumptions, especially for steelhead. In particular, the adult steelhead model assumes restricted spawning/rearing locations, and the steelhead models include no interactions with resident rainbow. The steelhead models relied more on information from outside the WVS because of a lack of steelhead data for the basin.

The utility function is used to summarize and compare results on a common scale and includes implicit assumptions about the relative importance of different population components in the analysis. The utility function is a simple proportional scoring of the model outputs on an arithmetic scale. In SDM, the choice of utility function should be a community decision based on considerations of the importance of objectives and criteria, so the use of a simplistic default function here makes the results hard to interpret in the context of objectives. At least some exploration of the sensitivity of results to other formulations of utility should be incorporated or addressed in a summary document. This was partially addressed for different weightings of the four models via “indifference curves” (Peterson et al. 2021), but not for alternative formulations of the utility functions.

b. Is the reasoning, rationale, and evidence for the key simplifying assumptions (e.g., no interactions with other species) reasonable and adequately documented?

Little or no rationale or evidence is presented to support the simplifying aspects of certain key assumptions. For example, the models assume complete ecological and reproductive separation between resident rainbow trout and steelhead. This is known to be untrue, so the real question is whether it is extensive enough to affect survival under differing flow alternatives and if there are adequate data from these populations to represent the relationship accurately. For example, populations across their range experience considerable gene flow between forms, and phenotypic plasticity affects which life history pattern the fish display. Whether the amount of gene flow between resident and anadromous *O. mykiss* is affected by hydrological flows and thereby influences steelhead survival is largely unknown. There may not be sufficient data to model this effect at present, but at least some acknowledgement of the weakness of the assumption and implications should be included. Depending on the river, the resident component may be a small or very large proportion of the species (numerically), affecting the scope for ecological interactions between the two forms.

3. Model validation and verification

- a. *Does the model documentation adequately describe testing steps utilized during model development (i.e., consistency check, sensitivity analyses, calibration, validation)?*

No specific model validation is described in the documents, but there is limited calibration to adult and smolt data for Chinook (no adequate steelhead data were available). The models are used in their generic formulation and are configured spatially for the Willamette system of rivers and dams, but the biological data used for parameter estimation is from multiple sources and systems. Some presentation of reality checks on the model results of population dynamics (even if information must be from other applications) is needed. The philosophy of the model relies on a bottom-up approach that presumes that if each component is grounded in empirical information, then the coupling of these will also generate realistic results. This is a valid approach but one that would benefit from some checks on the realism of modeling predictions. Each component is based on different years and types of data, so the legitimacy of combining them needs to be confirmed.

The Chinook models are calibrated to data on abundance at specific life stages (number of redds for adult Chinook, predicted adult equivalents for juvenile Chinook), but because of limited data for the WVS, there is no calibration for the steelhead models.

The reporting of erroneous parameter values in tables, incorrect citations, and issues of data relevance in the supplements to Peterson et al. (2021), some of which were resolved in correspondence between the modelers and ISAB, suggests that there may need to be more attention paid to the reporting of values used for the WVS application and concerning the confidence level appropriate for the limited site-specific validation. The authors may need to address these issues in future applications of the models as part of the SDM process.

- b. *Have the researchers adequately assessed the fit of model to current data? Have the researchers assessed the fit of the model to new data (e.g., use it to predict years not used in fitting and compare)?*

Of the four cohort models, the two Chinook salmon models were calibrated to abundance time series and demonstrated that the ranges of abundances produced by the models were generally comparable to those of the observations (Peterson et al. 2021, Supplement S3). No measures of fit were reported. For parameter estimates of the submodels, Peterson et al. (2021) presented R^2 or pseudo- R^2 as measures of the fit of the submodels. These appear to be reasonable large (typically 0.75) and are typically based on 2004-2018 data sets. The stream temperature and discharge models are based on 2001-2018 data. No assessment of fit to the additional years in the new data is described.

The analysis of PEIS alternatives (Peterson 2022b) used three specific years to typify the range of flow conditions. No assessment of model fit specific to those three years was reported.

- c. Has the model programming or system components been tested for computational correctness and associated errors? If not, what is the potential for errors to occur?*

While no code verification, QA/QC, or testing is described in the original papers, Peterson (2022a) stated that they verified predicted temperatures and flows against observations, which identified one submodel that was incorrectly coded. They also ran fish movement and distribution submodels individually to be sure they behaved as expected. They suggest that it “would be ideal to compare model predictions to future observations of adults spawning success and juvenile salmonid survival and movement.” Given the extensive use of the code and that it is in R so that others can access it, some simple testing would suffice. A relatively easy approach for checking code is to simulate each submodel for conditions with known correct responses. This does not evaluate the realism of the responses but rather that the code is correct.

- d. Has sensitivity and uncertainty analyses been applied to the model and the results used effectively to identify sensitive parameters and quantify variability around model predictions?*

Peterson et al. (2021) used sensitivity analysis to identify which components had the largest influence on estimated outcomes for the best candidate flow regime. Their approach (run one subcomponent from the smallest to the highest value based on the estimate and standard error) tends to identify the stage with the highest relative uncertainty, i.e., the stages where more data will lead to smaller uncertainty in the estimates. In addition, they used a similar response profile sensitivity analysis to assess the sensitivity of rankings of candidate flow actions to model components. An alternative way to identify sensitivity is to look, for example, at a 10% change in the estimate and see which component leads to a larger change.

4. Data gaps

- a. For which part of the model is data strongest and weakest? Is the model sensitive to these data gaps?*

Peterson et al. (2021) noted that the lack of steelhead data specific to the Willamette basin leads to low confidence in steelhead results. This issue is not unique to this model, but general to all the Willamette steelhead models. Here, they noted the lack of steelhead timing data, and the assumption of no interaction with resident rainbow trout and the assumptions regarding food supply are uncertain. Territory size is modeled to increase with fish size, which is reasonable. Larger fish need bigger territories, but as fish grow, others die, and so there may be

“habitat saturation” that is maintained if growth-dependent territory size matches mortality. However, this assumes that all fish are territorial. The submodels were fairly simple because typically less than 20 years of data are available. The calibration datasets for the decision-support models are only based on data from 2008-2017 which is a small data set for reliably fitting this type of model. See also specific comments above (Question 1.a) regarding data issues for the individual submodels such as the relationship between temperature and PSM, fecundity, and redd superimposition.

b. If so, then what type of research program is needed?

A research program focused on the WVS would improve data for habitat effects, spatial distributions and survivals, especially for steelhead, and would provide better inputs for this and other models of Willamette salmonids. For example, considering flood risk constraints, could a broader range of flows be explored and modeled for responses in the WVS? Also, given the stated importance of temperature profiles influencing en-route survival, some exploration of simulating and testing preferred thermal regimes would be helpful.

5. Model output

a. Can the outputs of the model be used to rigorously compare and contrast different management actions?

The OSU fish survival and flow models can be used to compare and contrast flow management alternatives, but the level of confidence associated with the comparisons is not quantified. The results have some close ties among alternatives and tradeoffs among species and life stages. As mentioned above, the utility function is somewhat subjective and discretionary as presented, as is the equal weighting of each decision-support model utility to get an overall score. The models are suitable for ranking alternative management actions under high, medium, and low-flow scenarios, but the use of their predictions as absolute responses in nature and predictions for specific years are questionable due to unknown uncertainties. Comparison with reality is also limited because of the reliance on multi-dimensional metrics without “natural” units (i.e., units that relate directly to observable quantities, in contrast to derived measures such as utilities, scores, or ranks). Dimensionless utility may be helpful for quick interpretations, but comparisons are also needed by output and in natural units, not only for this analysis as a standalone effort, but also when combined with the results of the other analyses. There is no confident and direct way to compare dimensionless, weighted-sum indicators across models. Utility scores are problematic without proper context. Table 1 in DeWeber and Peterson (2020)

is a good list of interpretable outcomes; Table 2 is a good example of challenges in interpretation with reliance on metrics not grounded in observable quantities.

There are many examples in economic analyses that merge various outputs. In this case, some decision will need to be made about the relative marginal value of an additional redd (output from Chinook adult model) or an additional adult equivalent (Chinook juvenile model). This was examined in the paper where different weightings were used (such as, the indifference curve analysis).

- i. How reliable are the comparisons of the effects of the alternatives on spring Chinook salmon and winter steelhead (e.g., can be used only to rank alternatives; can be used in cost-benefit analyses)?*

The models largely focus on physical (i.e., hydrological, temperature) outcomes to deduce biological ones. It is unclear how robust the deductions/interpretations/decisions are to errors and variability in the physical model outputs. DeWeber and Peterson (2020, p. 609) focus on interannual variability in water availability and air temperature; however, uncertainty is not available for the ResSim model. Thus, it is not feasible with the present information to quantify uncertainties due to uncertainty in the model inputs generated from the other models (ResSim and CE-QUAL-W2).

Using a formal optimization algorithm is an excellent way to identify good and bad flows patterns. Optimization was used by Peterson et al. (2021) in two ways. First, the models were used to identify the flow pattern that maximizes summary measures of the four model output variables. Second, they were used to model output variables and summary measures for seven specified flow management alternatives to identify which alternative maximizes the measure. In the analysis of PEIS alternatives (Peterson 2022b), there were two outputs for Chinook and two outputs for steelhead: (1) Chinook: surviving redds (adult submodel), (2) Chinook: adult equivalents (juvenile submodel), (3) Steelhead: age 1 juveniles (adult submodel), and (4) Steelhead: smolts surviving to Willamette Falls (smolt submodel). There were also two ways to combine the two outputs per species using utilities: equal weights additive and multiplied, and three (narrow, median, broad) assumed habitat levels. Seven alternative management scenarios (including NAA) were simulated. For each output, the results were reported in natural units, utility, and as relative loss.

The only use of the optimization of the flows was in Peterson et al. (2021) to provide the optimal value for computing relative loss. Other than that, the “optimization” was to use the model to generate the outputs for each management scenario in order to allow for comparison and identification of which scenarios generated outputs, higher utility, or lower relative loss. There was no searching of alternative flow patterns – the seven alternative flow patterns were

given to the models. The models summarized their performance, with one summary measure using the optimized flow pattern (not one of the seven).

The reporting of results by tributary (these are outcomes for a scenario) appears to be very informative.

The various ways used to summarize the performance of each of the seven flow management alternatives can be confusing. Use of ranks and utility can create answers that rely on exaggerated differences among the alternatives. Indeed, the purpose of utilities is to highlight differences by rescaling original units into zero-to-one regardless of how different the output was across alternatives. The more relevant outputs for decision-making in this situation would seem to emphasize the outputs in their natural units and relative loss.

ii. Is the output stochastic so questions about quasi-extinction can be answered?

Because these are not complete life-cycle models, they cannot be used to directly address questions of extinction risk.

iii. Do the researchers distinguish whether differences in predicted responses among alternatives are biologically meaningful?

Yes, the team does provide some interpretations, which are based on biology. The predicted responses are within the ranges of recent and recorded conditions. One key uncertainty will be the net temperature and precipitation effects under emerging climatic conditions (see 5c below).

b. What specific metrics or types of output are provided to the user? Are these outputs sufficient for evaluating the EIS alternatives? Would additional types of output be useful?

Output metrics are Chinook surviving redds, Chinook adult equivalent returns, steelhead age-1 juveniles, and steelhead smolts (Peterson 2022b, Table 1). They relate to the two PEIS survival metrics and are not intended to be sufficient for fully evaluating EIS alternatives, but they can provide relevant information to the evaluation. In general, the model outputs are appropriate for the goals of the analysis and are presented in an understandable way. Some indication of confidence (statistical and otherwise) in the results would be useful. As well, the submodel parts are “hidden” from the user and what the user sees for a given water input is the final value from the decision-support model without understanding of the processes and uncertainties that contribute to the results. The graphs in Peterson et al. (2021) showing the relative utility are useful but fail to show the variability in utility over the simulations. For example, two flow regimes with the same mean utility may have quite different ranges of utilities in the simulations and the one with a smaller variability may be preferred. Despite that

limitation, similar graphs would be useful in comparing the EIS alternatives, rather than the plain tables in the EIS analysis results (Peterson 2022b).

The submodel outputs and relative loss are very informative, whereas the summary metrics based on utility and on rankings are easily misinterpreted.

In general, the results suggest relatively minor differences at the basin-scale among the management alternatives. Table 2 shows basinwide outputs in natural units averaged for the three years, Table 4 shows outputs as relative losses averaged for the three years, and Tables 5 and 7 show Table 2 and Table 4 by tributary but for 2015 only (not averaged over the three years). There were larger and ecologically significant differences for certain tributaries across management alternatives. It is not discussed whether the underlying models have been assessed for skill at the tributary level. Results aggregated for the basin-level average out many of these tributary-level differences. Then, results are further distorted by using utilities and rankings.

The results suggest that there is no global optimal alternative among the seven flow management alternatives. The authors state in the Draft Report (Peterson 2022b): “The simulations indicated that, with the exception of Steelhead trout smolt survival, no single flow alternative was best across all years and tributaries (Table 1).” The authors further state: “The utilities calculated using the basin-wide outcomes averaged across years indicated that the effects of the flow management alternatives varied by species and life history stage (Table 3).” Therefore, one could conclude that selection among the alternatives should be based on other considerations than an overall (basinwide with both species combined) effect. Perhaps looking at the results by species or by tributary. Further, there are some situations of relatively large relative loss, which suggests that perhaps other alternative scenarios should be explored. Large relative loss indicates there is room for improvement (because it is relative to an optimal flow pattern).

- c. *How far outside of the envelope of data used to develop the model will climate change take us? Do we have good information on how the model will perform outside of its development based on historical conditions?*

No information is presented on impacts of climate change or regarding model performance outside the range of data for which the models were calibrated. However, given that regional climate predictions are expected to greatly alter flow and temperature regimes over the next several decades and beyond, this kind of information/simulation is warranted.

6. Model improvement

a. What improvements can be made to address any critical shortcomings of the models?

The performance of the models under climate change is unknown. For each decision-support model and each submodel, it is necessary to see if the range of data supported by the submodel encompasses likely scenarios under climate change.

The decision-support results are merged using a utility function that is difficult to understand with regard to its application to guiding decisions on flow regimes. It is unclear how this utility function responds to selective inclusion of various scenarios, e.g., what is the impact of including scenarios known to perform badly and so make the “worst” value smaller for some of the decision-support models. For example, suppose that 99% of the results were similar but the “worst” value was so small, that the utility of the decision model will essentially now be 1 because of the one “worst” value. More consideration to the structure and weightings of the utility function is needed.

b. Where should future model development be focused?

The first step in improving this approach should focus on using the sensitivity analysis results to prioritize research and monitoring to improve data for those components that strongly influence results. Additional stochasticity (demographic, environmental) could be included both to better represent uncertainty and (potentially) to evaluate extinction risk. Consideration should also be given to some of the simplifying assumptions such as ecology/gene flow and whether it alters output and interpretation (especially flow and temperature effects). Finally, the model documentation would benefit from some clarification of issues regarding parameters in tables, citations, and data relevance identified under Question 1.a above.

Appendix 1: ISAB Review Questions for the Individual Models and the Multi-model Approach

To review the models, the ISAB developed a detailed set of questions, which integrated ten questions in the [Corps 2021 review request memorandum](#), two follow-up questions provided by the Corps, the ISAB's 2014 review of NOAA's Willamette model, and internal ISAB discussions.

A. Questions for Individual Models

1. Model structure and development

- a. Is the model structure appropriate for evaluating fish responses for the EIS alternatives?
 - i. Are the relevant processes (mortality, growth, reproduction, movement) included and are they represented in a way that encompasses the changes within and among the alternatives?
 - ii. Are the key functional forms (e.g., shapes, dependence on hydrological and environmental variables) sufficiently realistic?
 - iii. What aspects of the model are stochastic?
- b. Are the models formulated to assess fish responses at the appropriate temporal and spatial scales that both capture the effects of the alternatives and generate outcomes relevant to management scales?
- c. Are the estimates used for model input parameters (hydrological, environmental, processes) reasonable and scientifically defensible? What data are used to estimate or confirm model parameter values?
- d. What sources of variation that affect population dynamics and responses have been accounted for?
- e. Have the researchers adequately evaluated the sensitivity of the predictions to uncertainties in various sources of data?
- f. How sensitive are predictions to uncertainties in hydrological conditions, environmental variation, and model parameters?
- g. Are subcomponent models adequately integrated into the larger model?

2. Key assumptions and limitations

- a. Are the model assumptions and limitations sufficiently documented? What are the key assumptions, strengths, and limitations?
- b. Are the reasoning, rationale, and evidence for the key simplifying assumptions (e.g., no interactions with other species) reasonable and adequately documented?

3. Model validation and verification

- a. Does the model documentation adequately describe testing steps utilized during model development (i.e., consistency check, sensitivity analyses, calibration, validation)?
- b. Have the researchers adequately assessed the fit of model to current data? Have the researchers assessed the fit of the model to new data (e.g., use it to predict years not used in fitting and compare)?
- c. Have the model programming or system components been tested for computational correctness and associated errors? If not, what is the potential for errors to occur?
- d. Have sensitivity and uncertainty analyses been applied to the model and the results used effectively to identify sensitive parameters and quantify variability around model predictions?

4. Data gaps

- a. For which part of the model is data strongest and weakest? Is the model sensitive to these data gaps?
- b. If so, then what type of research program is needed?

5. Model output

- a. Do the models support effective estimation of the metrics for comparisons of effects of the alternatives and measures on spring Chinook salmon and winter steelhead (Table 1), considering the types and range of measures and alternatives being evaluated in the WVS EIS?

Table 1. Metrics for comparisons of effects of the alternatives and measures on spring Chinook salmon and winter steelhead.

| <i>Metric</i> | <i>Description</i> | <i>Units</i> |
|---------------------------------------|--|--|
| Spatial Structure | Habitat Quality, Spatial Configuration, Dispersal dynamics of a population | Estimate of change in spatial and temporal availability of habitat; spatial and temporal distribution of natural spawners |
| Diversity | Diversity in juvenile rearing patterns; genetic diversity and influence from hatchery spawners | Estimate of change in timing and size when juveniles emigrating past each dam and to reaches downstream (e.g. Willamette Falls), their relative adult return rates; genetic influence of hatcheries as measured by pHOS and pNOB; effective population size. |
| Productivity | Population Growth Rate | Estimated R/S, SAR, Estimated fry to "smolt" survival passing dams |
| Abundance | Geomean of naturally producing adults | Estimated trends in geomean abundance of natural origin spawners over several salmon generations (forecasted period will depend on data available to support given expected uncertainty of prospective analyses). |
| Extinction Risk | Quasi extinction risk (QET) | Probability that abundance is less than threshold over several salmon generations (forecasted period will depend on data available to support given expected uncertainty of prospective analyses). |
| Dam passage juvenile survival | Dam passage survival of fry, sub-yearling and yearling juveniles | Estimated annual juvenile passage survival rate at each dam by lifestage (three passage age groups – e.g. "fry", "sub-yearlings", and "yearlings") |
| WVS dam discharge effects on survival | Juvenile survival, adult equivalents, pre-spawn mortality | Estimate of change in timing and size when juveniles emigrating past each dam and to reaches downstream (e.g. Willamette Falls), their relative adult return rates; genetic influence of hatcheries as measured by pHOS and pNOB; effective population size. |

- b. Can the outputs of the model be used to rigorously compare and contrast different management actions?
 - i. How reliable are the comparisons of the effects of the alternatives on spring Chinook salmon and winter steelhead (e.g., can be used only to rank alternatives; can be used in cost-benefit analyses)?
 - ii. Is the output stochastic so questions about quasi-extinction can be answered?
 - iii. Do the researchers distinguish whether differences in predicted responses among alternatives are biologically meaningful?
 - c. What specific metrics or types of output are provided to the user? Are these outputs sufficient for evaluating the EIS alternatives? Would additional types of output be useful?
 - d. How far outside of the envelope of data used to develop the model will climate change take us? Do we have good information on how the model will perform outside of its development based on historical conditions?
- 6. Model improvement**
- a. What improvements can be made to address any critical shortcomings of the models?
 - b. Where should future model development be focused?

B. Questions for the ISAB's recommendations to the Corps about using multiple models

1. Evaluation of management strategies

Is the evaluation process consistent with best practices for management strategy evaluation (Punt et al. 2016)?

2. Model coupling

Are models coupled (e.g., directly, by passing files)?

3. Model components

Is this model used as a subcomponent of a larger model? How well is it integrated into the larger model? How critical is this sub-model to the larger model?

4. Information representation

- a. Are the aggregation/disaggregation schemes adequate for matching the common inputs used by two or more models (e.g., ResSim output)?
- b. Are the aggregation/disaggregation schemes adequate for matching the output of the donor model to the scales of the receiving model?

5. Model comparisons

- a. Are the outputs of the multiple models used consistently and effectively to compare and contrast different management actions?
- b. How reliable are the comparisons (e.g., only use to rank alternatives; can they be used in cost-benefit analyses)?
- c. Are the metrics of fish responses from the different models consistent and adequate for comparing alternatives (e.g., are the outputs stochastic or deterministic so questions about quasi-extinction can be answered)?
- d. How consistent are process representations (e.g., growth, movement, passage effects) that occur in more than one of the models?

6. Using the models after the completion of the EIS

If any of the models are used beyond the EIS analyses, what are the plans for long-term management of the model(s) development?

- a. Is there a version control system?
- b. Are updates done continuously or in batch increments?
- c. How are “bugs” fixed?
- d. Are model versions, submodels, and input data documented and versions tracked?
- e. Does a steering committee coordinate the long-term management of the models?

7. Communication to managers and stakeholders

- a. Does the USACE have a reasonable strategy for obtaining input from managers, federal and state cooperators, and stakeholders about the models and results of evaluation alternatives prior to completing the EIS?
- b. Does the USACE have a reasonable strategy and operational approach for communicating the results of future applications of the multiple models to managers, federal and state cooperators, and stakeholders after the completion of the EIS?

References

- Anderson, J.H., and P.C. Topping. 2018. Juvenile life history diversity and freshwater productivity of Chinook Salmon in the Green River, Washington. *North American Journal of Fisheries Management* 38:180-193.a
- Barnett, H.K., T.P. Quinn, M. Bhuthimethee, and J.R. Winton. 2020. Increased prespawning mortality threatens an integrated natural- and hatchery-origin sockeye salmon population in the Lake Washington Basin. *Fisheries Research* 227:105527.
- Beamish, R.J., R.M. Sweeting, C.C. Neville, K.L. Lange, T.D. Beacham, and D. Preikshot. 2012. Wild chinook salmon survive better than hatchery salmon in a period of poor production. *Environmental Biology of Fishes* 94:135–148.
- Blair, G.R., L.C. Lestelle, and L.E. Mobrand. 2009. The Ecosystem Diagnosis and Treatment Model: A Tool for Assessing Salmonid Performance Potential Based on Habitat Conditions. E.E. Knudsen, editor, *Pacific Salmon Environment and Life History Models*. Bethesda, MD: American Fisheries Society; pp. 289–309.
- Bonan, G.B., and S.C. Doney. 2018. Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science* 359.6375: eaam8328.
- Bowerman, T., A. Roumasset, M.L. Keefer, C.S. Sharpe, and C.C. Caudill. 2018. Prespawn mortality of female Chinook Salmon increases with water temperature and percent hatchery origin. *Transactions of the American Fisheries Society* 147: 31-42.
<https://doi.org/10.1002/tafs.10022>
- Bowerman, T.E., M.L. Keefer, and C.C. Caudill. 2021. Elevated stream temperature, origin, and individual size influence Chinook salmon prespawn mortality across the Columbia River Basin. *Fisheries Research* 237:105874.
- Bradford, M.J., J.A. Grout, and S. Moodie. 2001. Ecology of juvenile Chinook salmon in a small non-natal stream of the Yukon River drainage and the role of ice conditions on their distribution and survival. *Canadian Journal of Fisheries and Aquatic Sciences* 79:2043-2054.
- Bulkley, R.V. 1967. Fecundity of steelhead trout, *Salmo gairdneri*, from Alsea River, Oregon. *Journal of the Fisheries Research Board of Canada* 24(5):917-926.
- Burke, B.J., W.T. Peterson, B.R. Beckman, C. Morgan, E.A. Daly, and M. Litz. 2013. Multivariate models of adult Pacific salmon returns. *PLoS ONE* 8, e54134 (doi:10.1371/journal.pone.0054134).

- Chagaris, D., S. Sagarese, N. Farmer, B. Mahmoudi, K. de Mutsert, S. VanderKooy, W.F. Patterson III, M. Kilgour, A. Schueller, R. Ahrens, and M. Lauretta. 2019. Management challenges are opportunities for fisheries ecosystem models in the Gulf of Mexico. *Marine Policy* 101: 1-7.
- Chasco B., B. Burke, L. Crozier, and R. Zabel. 2021. Differential impacts of freshwater and marine covariates on wild and hatchery Chinook salmon marine survival. *PLoS ONE* 16(2): e0246659. <https://doi.org/10.1371/journal.pone.0246659>
- Conroy, M.J., R.J. Barker, P.J. Dillingham, D. Fletcher, A.M. Gormley, and I.M. Westbrooke. 2008. Application of decision theory to conservation management: recovery of Hector's dolphin. *Wildlife Research* 35:93-102.
- Cramer, S.P. 2001. The relationship of stream habitat features to potential for production of four salmonid species. Final report to the Oregon Building Industry Association. S. P. Cramer and Associates, Gresham, Oregon.
- Daum, D.W., and B.G. Flannery. 2011. Canadian-origin Chinook salmon rearing in nonnatal U.S. tributary streams of the Yukon River, Alaska. *Transactions of the American Fisheries Society* 140:207-220.
- DeWeber, J.T., and J.T. Peterson. 2020. Comparing environmental flow implementation options with structured decision making: case study from the Willamette River, Oregon. *Journal of the American Water Resources Association* 56: 599-614.
- Firman, J., R. Schroeder, R. Lindsay, K. Kenaston, and M. Hogansen. 2003. Work completed for compliance with the Biological Opinion for hatchery programs in the Willamette Basin, USACE funding: 2003. Oregon Department of Fish and Wildlife, Task Order: NWP-OP-FH-02-01, Salem.
- Fleishman, E., editor. 2023. Sixth Oregon climate assessment. Oregon Climate Change Research Institute, Oregon State University, Corvallis, Oregon. DOI: 10.5399/osu/1161.
- Ford, M.J., editor. 2022. Biological Viability Assessment Update for Pacific Salmon and Steelhead Listed Under the Endangered Species Act: Pacific Northwest. Northwest Fisheries Science Center (U.S.).doi:10.25923/KQ2N-KE70.
- Fordham, D.A., T.M. Wigley, M.J. Watts, and B.W. Brook. 2012. Strengthening forecasts of climate change impacts with multi-model ensemble averaged projections using MAGICC/SCENGEN 5.3. *Ecography* 35(1): 4-8.
- Gimenez, O., B.J. Morgan, and S.P. Brooks. 2009. Weak Identifiability in Models for Mark-Recapture-Recovery Data. In: Thomson, D.L., Cooch, E.G., Conroy, M.J. (eds) *Modeling Demographic Processes In Marked Populations*. Environmental and Ecological Statistics, vol 3. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-78151-8_48

- Gregory, R., L. Failing, M. Harstone, G. Long, T. McDaniels, and D. Ohlson. 2012. Structured decision making: a practical guide to environmental management choices. John Wiley & Sons, Chichester.
- Hamill, T.M., M.J. Brennan, B. Brown, M. DeMaria, E.N. Rappaport, and Z. Toth. 2012. NOAA's future ensemble-based hurricane forecast products. *Bulletin of the American Meteorological Society* 93(2): 209-220.
- Hersbach, H., C. Peubey, A. Simmons, P. Berrisford, P. Poli, and D. Dee. 2015. ERA-20CM: A twentieth-century atmospheric model ensemble. *Quarterly Journal of the Royal Meteorological Society* 141(691): 2350-2375.
- Hill, G., S. Kolmes, M. Humphreys, R. McLain, and E.T. Jones. 2019. Using decision support tools in multistakeholder environmental planning: restorative justice and subbasin planning in the Columbia River Basin. *Journal of Environmental Studies and Sciences* 9:170-186. DOI:10.1007/s13412-019-00548-x
- ICF. 2022. Ecosystem Diagnosis & Treatment to Support the Willamette Valley Systems Environmental Impact Statement. Final. May. (ICF IWO191.0.195.01) Portland, OR. Prepared for USACE-Portland District, Portland, OR.
- Independent Scientific Advisory Board (ISAB). 2014-3. Review of the Fish Benefits Workbook for the U.S. Army Corps of Engineers Willamette Valley Project. Northwest Power and Conservation Council, Portland, Oregon ([ISAB 2014-3](#)).
- ISAB. 2014-4. Review of NOAA Fisheries' Viable Salmonid Population (VSP) Modeling of Willamette River Spring Chinook Populations. Northwest Power and Conservation Council, Portland, Oregon ([ISAB 2014-4](#)).
- ISAB. 2017-1. Review of NOAA Fisheries' Interior Columbia Basin Life-Cycle Modeling Draft Report. Northwest Power and Conservation Council, Portland, Oregon ([ISAB 2017-1](#)).
- ISAB's Review of the Upper Columbia United Tribes' Fish Passage and Reintroduction Phase 1 Report: Investigations Upstream of Chief Joseph and Grand Coulee Dams. Northwest Power and Conservation Council, Portland, Oregon ([ISAB 2019-3](#))
- ISAB. 2022-2. Review of the Upper Columbia United Tribes' Phase 2 Implementation Plan: Testing Feasibility of Reintroduced Salmon in the Upper Columbia River Basin. Northwest Power and Conservation Council, Portland, Oregon ([ISAB 2022-2](#)).
- ISAB/ISRP. 2004-13. Scientific Review of Subbasin Plans for the Columbia River Basin Fish and Wildlife Program. Northwest Power and Conservation Council, Portland, Oregon ([ISAB 2004-13](#))

- Independent Scientific Review Panel (ISRP). 2011-26. Review of the Research, Monitoring, and Evaluation Plan and Proposals for the Willamette Valley Project. Northwest Power and Conservation Council, Portland, Oregon ([ISRP 2011-26](#)).
- Jardim, E., M. Azevedo, J. Brodziak, E.N. Brooks, K.F. Johnson, N. Klibansky, C.P. Millar, C. Minto, I. Mosqueira, R.D. Nash, and P. Vasilakopoulos. 2021. Operationalizing ensemble models for scientific advice to fisheries management. *ICES Journal of Marine Science* 78(4): 1209-1216.
- Klipsch, J.D., G.C. Modini, S.M. O'Connell, M.B. Hurst, and D.L. Black. 2021. HEC-ResSim reservoir system simulation User's manual, version 3.3: U.S. Army Corps of Engineers hydrologic engineering center report CPD-82, 670 p, Accessed June 11, 2021 at https://www.hec.usace.army.mil/software/hec-ressim/documentation/HEC-ResSim_33_UsersManual.pdf.
- Lestelle, L.C., J.A. Lichatowich, L.E. Mobrand, and V.I. Cullinan. 1994. Ecosystem diagnosis and treatment planning model as applied to supplementation: model description, user guide, and theoretical documentation for the model introduced in the Summary Report Series. Report to Bonneville Power Administration, Project Number 92-018, March 1994.
- Lestelle, L.C., L.E. Mobrand, and W.E. McConnaha. 2004. Information structure of Ecosystem Diagnosis and Treatment (EDT) and Habitat Rating Rules for Chinook salmon, coho salmon, and steelhead trout. Mobrand Biometrics, Vashon Island, Washington.
- Lestelle, L.C. 2005. Guidelines for rating level 2 environmental attributes in Ecosystem Diagnosis and Treatment (EDT). Jones & Stokes Associates, Inc.
- Lewis, K.A., K.A. Rose, K. De Mutsert, S. Sable, C. Ainsworth, D.C. Brady, and H. Townsend. 2021. Using multiple ecological models to inform environmental decision-making. *Frontiers in Marine Science* 283.
- Lin, J., K. Emanuel, and J.L. Vigh. 2020. Forecasts of hurricanes using large-ensemble outputs. *Weather and Forecasting* 35(5): 1713-1731.
- Litzow, M.A., L. Ciannelli, P. Puerta, J.J. Wettstein, R.R. Rykaczewski, and M. Opiekun. 2018. Non-stationary climate–salmon relationships in the Gulf of Alaska. *Proceedings of the Royal Society B: Biological Sciences* 285, 20181855.
- McElhany, P., M.H. Ruckelshaus, M.J. Ford, T.C. Wainwright, and E.P. Bjorkstedt. 2000. Viable salmonid populations and the recovery of evolutionarily significant units. U.S. Dept. Commerce, NOAA Tech. Memo. NMFS-NWFSC-42, 156 p.
- McElhany, P., C. Busack, M.W. Chilcote, S. Kolmes, B.A. McIntosh, J. Myers, D. Rawding, E.A. Steel, C. Steward, D.L. Ward, T. Whitesel, and C. Willis. 2006. Revised viability criteria for salmon and steelhead in the Willamette and Lower Columbia Basins. Willamette/Lower Columbia Technical Recovery Team and Oregon Department of Fish and Wildlife. 178 p.

- McElhany, P., E.A. Steel, K. Avery, N. Yoder, C. Busack, and B. Thompson. 2010. Dealing with uncertainty in ecosystem models: lessons from a complex salmon model. *Ecological Applications* 20:465–482.
- Murray, C.B., and J.D. McPhail. 1988. Effect of incubation temperature on the development of five species of Pacific salmon (*Oncorhynchus*) embryos and alevins. *Canadian Journal of Zoology* 66:266-273.
- Murray, C.B., and M.L. Rosenau. 1989. Rearing of juvenile chinook salmon in nonnatal tributaries of the lower Fraser River, British Columbia. *Transactions of the American Fisheries Society* 118:284-289.
- Nisbet, M.C., and D.A. Scheufele. 2009. What's next for science communication? Promising directions and lingering distractions. *American Journal of Botany* 96(10): 1767-1778.
- Parker, W.S. 2013. Ensemble modeling, uncertainty and robust predictions. *Wiley Interdisciplinary Reviews: Climate Change* 4(3): 213-223.
- Peterman, R.M. 2004. Possible solutions to some challenges facing fisheries scientists and managers. *ICES Journal of Marine Science* 61(8): 1331-1343.
- Peterson, J.T., J.E. Pease, L. Whitman, J. White, L. Stratton-Garvin, S. Rounds, and R. Wallick. 2021. Integrated tools for identifying optimal flow regimes and evaluating alternative minimum flows for recovering at-risk salmonids in a highly managed system. *River Research and Applications* 38(2):293-308.
- Peterson, J.T. 2022a. Response to “Request to USACE for Willamette Review by ISAB.” E-mail from J. Peterson to E. Merrill, 4 April 2022.
- Peterson, J.T. 2022b. Evaluation of the effectiveness of alternative Willamette River flow regimes for supporting at-risk salmonids: Draft report for U.S. Army Corps of Engineers, Portland District. Oregon Cooperative Fish and Wildlife Research Unit, Oregon State University. Corvallis, OR. 25 p.
- Phillis, C.C., A.M. Sturrock, R.C. Johnson, and P.K. Weber. 2018. Endangered winter-run Chinook salmon rely on diverse rearing habitats in a highly altered landscape. *Biological Conservation* 217:358-362.
- Quinn, T.P. 2018. *The behavior and ecology of Pacific salmon and trout*, 2nd ed. American Fisheries Society, Bethesda, Maryland.
- Railsback, S.F. 2022. What we don't know about the effects of temperature on salmonid growth. *Transactions of the American Fisheries* 151:3-12.

- Ralls, K., S.R. Beissinger, and J.F. Cochrane. 2002. Guidelines for using population viability analysis in endangered-species management, in: Beissinger, S.R., McCullough, D.R. (Eds.), *Population Viability Analysis*. University of Chicago Press, Chicago, pp. 521–550.
- Rawding, D. 2004. Comparison of spawner-recruit data with estimates of EDT spawner-recruit performance. Obtained from Appendix E. Washington Lower Columbia Salmon Recovery and Fish and Wildlife Subbasin Plans, May 2010.
- Raymond, H. L. 1988. Effects of hydroelectric development and fisheries enhancement on spring and summer chinook salmon and steelhead in the Columbia River Basin. *North American Journal of Fisheries Management* 8:1-24.
- Refsgaard, J.C., J.P. Van der Sluijs, J. Brown, and P. Van der Keur. 2006. A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources* 29(11): 1586-1597.
- Rindskopf, D. 2002. Infinite Parameter Estimates in Logistic Regression: Opportunities, Not Problems. *Journal of Educational and Behavioral Statistics* 27:147–61.
- Roni P., P. Anders, T.J. Beechie, and D.J. Kaplowe. 2018. Review of tools for identifying, planning, and implementing habitat restoration for Pacific Salmon and Steelhead. *North American Journal of Fisheries Management* 38: 355-376. 10.1002/nafm.10035
- Rose, K.A., R.B. Cook, A.L. Brenkert, R.H. Gardner, and J.P. Hettelingh. 1991a. Systematic comparison of ILWAS, MAGIC, and ETD watershed acidification models: 1. Mapping among model inputs and deterministic results. *Water Resources Research* 27(10): 2577-2589.
- Rose, K.A., A.L. Brenkert, R.B. Cook, R.H. Gardner, and J.P. Hettelingh. 1991b. Systematic comparison of ILWAS, MAGIC, and ETD watershed acidification models: 2. Monte Carlo analysis under regional variability. *Water Resources Research* 27(10): 2591-2603.
- Rose, K.A., and S.E. Sable. 2013. The 2017 Coastal Master Plan: Strategy for Selecting Fish Modeling Approaches. Coastal Protection and Restoration Authority, Baton Rouge, LA, 122 pp.
- Rupp, D.E., T.C. Wainwright, P.W. Lawson, and W.T. Peterson. 2011. Marine environment-based forecasting of coho salmon (*Oncorhynchus kisutch*) adult recruitment. *Fisheries Oceanography* 21, 1-19 [+ Errata vol. 21:226-227].
- Schmolke, A., P. Thorbek, D.L. DeAngelis, and V. Grimm. 2010. Ecological models supporting environmental decision making: a strategy for the future. *Trends in Ecology & Evolution* 25(8): 479-486.
- Scrivener, J.C., T.G. Brown, and B.C. Anderson. 1994. Juvenile salmon (*Oncorhynchus tshawytscha*) utilization of Hawks Creek, a small and nonnatal tributary of the upper Fraser River. *Canadian Journal of Fisheries and Aquatic Sciences* 51:1139-1146.

- Steel, E.A., P. McElhany, N.J. Yoder, M.D. Purser, K. Malone, B.E. Thompson, K.A. Avery, D. Jensen, G. Blair, C. Busack, M.D. Bowen, J. Hubble, and T. Kantz. 2009. Making the best use of modeled data: multiple approaches to sensitivity analysis of a fish-habitat model. *Fisheries* 34:330–339.
- Sullivan, K., D.J. Martin, R.D. Cardwell, J.E. Toll, and S. Duke. 2000. An analysis of the effects of temperature on salmonids of the Pacific Northwest with implications for selecting temperature criteria. Sustainable Ecosystems Institute, Portland, Oregon.
- Thompson, E. 2022. *Escape from model land*. Basic Books, New York.
- U.S. Army Corps of Engineers (USACE). 2019. The Willamette Valley Project. U.S. Army Corps of Engineers, Portland District, Portland, OR. 2 p. (Avail. <https://usace.contentdm.oclc.org/utis/getfile/collection/p16021coll6/id/2151>, accessed 2023-03-03).
- U.S. Army Corps of Engineers (USACE). 2022. The Willamette Valley System Operations and Maintenance Draft Programmatic Environmental Impact Statement. U.S. Army Corps of Engineers, Portland District, Portland, OR. (Avail. <https://www.nwp.usace.army.mil/Locations/Willamette-Valley/System-Evaluation-EIS/>, accessed 2023-03-03).
- van der Sleen, P., P.A. Zuidema, J. Morrongiello, J.L. Ong, R.R. Rykaczewski, W.J. Sydeman, E. Di Lorenzo, and B.A. Black. 2022. Interannual temperature variability is a principal driver of low-frequency fluctuations in marine fish populations. *Communications Biology* 5: 1-8. <https://doi.org/10.1038/s42003-021-02960-y>
- Wainwright, T.C., 2021. Ephemeral relationships in salmon forecasting: A cautionary tale. *Progress in Oceanography* 193, 102522.
- Welch, D.W., A.D. Porter, and E.L. Rechisky. 2021. A synthesis of the coast-wide decline in survival of West Coast Chinook Salmon (*Oncorhynchus tshawytscha*, Salmonidae). *Fish and Fisheries* 22: 194– 211. <https://doi.org/10.1111/faf.12514>
- Wells, S. 2020. CE-QUAL-W2: A two-dimensional, laterally averaged, hydrodynamic and water quality model, Version 4.5. Department of Civil and Environmental Engineering, Portland State University, Portland, Oregon. <http://www.ce.pdx.edu/w2/>
- Welp, M., A. de la Vega-Leinert, S. Stoll-Kleemann, and C.C. Jaeger. 2006. Science-based stakeholder dialogues: Theories and tools. *Global Environmental Change* 16(2): 170-181.
- Westley, P.A.H., T.P. Quinn, and A.H. Dittman. 2013. Rates of straying by hatchery-produced Pacific salmon (*Oncorhynchus* spp.) and steelhead (*Oncorhynchus mykiss*) differ among species, life history types, and populations. *Canadian Journal of Fisheries and Aquatic Sciences* 70:735-746.

- Woodson, L.E., B.K. Wells, P.K. Weber, R.B. MacFarlane, G.E. Whitman, and R.C. Johnson. 2013. Size, growth, and origin-dependent mortality of juvenile Chinook salmon *Oncorhynchus tshawytscha* during early ocean residence. *Marine Ecology Progress Series* 487:163–175.
- Yoder, N.J., M.D. Bowen, K.P. Zehfuss, J. Hubble, B. Watson, and G.R. Blair. 2006. Systematic sensitivity analysis of the EDT model applied to selected fish populations in Yakima River basin. U.S. Bureau of Reclamation Technical Memorandum 86-68290-11, Denver, CO.
- Zabel, R., J. Myers, R. Chittaro, and J. Jorgensen. 2015. Viable Salmonid Population (VSP) modeling of Willamette River spring Chinook and steelhead populations. Northwest Fisheries Science Center, National Oceanic and Atmospheric Administration, Seattle.
- Zhou, S. 2002. Size-dependent recovery of chinook salmon in carcass surveys. *Transactions of the American Fisheries Society* 131:1194-1202.
- Zimmerman, M.S., C. Kinsel, E. Beamer, E.J. Connor, and D.E. Pflug. 2015. Abundance, survival, and life history strategies of juvenile Chinook Salmon in the Skagit River, Washington. *Transactions of the American Fisheries Society* 144:627-641.